Deciphering the Molecular Underpinnings of the Cryptic Cellobiose Metabolism in *Escherichia coli*

(An omics guided approach to modularize CBP microbes)

Vinuselvi Parisutham

School of Life Sciences

(Department of Biomedical Engineering)

Graduate School of UNIST

2014

Deciphering the Molecular Underpinnings of the Cryptic Cellobiose Metabolism in *Escherichia coli*

(An omics guided approach to modularize CBP microbes)

Vinuselvi Parisutham

School of Life Sciences

(Department of Biomedical Engineering)

Graduate School of UNIST

2014

Deciphering the Molecular Underpinnings of the Cryptic Cellobiose Metabolism in *Escherichia coli*

(An omics guided approach to modularize CBP microbes)

A thesis submitted to the Graduate School of UNIST in partial fulfillment of the requirements for the degree of Combined-Master of Science and Doctor of Philosophy

Vinuselvi Parisutham

12.15.2014

Approved by

Advisor
Prof. Sung Kuk Lee

Deciphering the Molecular Underpinnings of the Cryptic Cellobiose Metabolism in *Escherichia coli*

(An omics guided approach to modularize CBP microbes)

Vinuselvi Parisutham

This certifies that the thesis of Vinuselvi Parisutham is approved.

12. 15. 2014

Signature

_____

Advisor: Prof. Sung Kuk Lee

Signature

_____

Prof. Gyoo Yeol Jung

Signature

_____

Prof. Robert James Mitchell

Signature

_____

Prof. Cheol-Min Ghim

Signature

_____

Prof. Dougu Nam

Deciphering the Molecular Underpinnings of the Cryptic Cellobiose Metabolism in *Escherichia coli*

(An omics guided approach to modularize CBP microbes)

Summary

Consolidated bioprocessing (CBP) is an efficient process that combines saccharification and fermentation of lignocellulosic biomass into a single microbial host. The choice of ideal microbes for CBP is challenging as it demands efficient functioning of several complex traits including expression of sets of saccharifying enzymes, metabolism of wide range of substrates, tolerance to various inhibitors, and high yield of desired products. One approach to develop an ideal host for CBP is to mimic the native cellulolytic microbes. In accordance with this, most native cellulolytic microbes metabolize cellulose in the form of cellobiose to obtain energetic benefits for growth on cellulose and to avoid feed-back inhibition of cellulase by glucose or cellobiose. In this study, we constructed cellobiose-metabolizing *Escherichia coli* (named as strain OSS, Original Synthetic Strain or ESS, Evolved Synthetic Strain which differ in their ability to ferment cellobiose) by exploiting its native cryptic *chb* and *asc* operons with an aim for using it as a platform host for Consolidated bioprocessing (CBP) or Simultaneous Saccharification and Fermentation (SSF) process. In addition to paving way for efficient consolidated bioprocessing, in depth analysis of strain ESS revealed several interesting molecular mysteries and regulations related to the cryptic operons of *E. coli*. Noteworthy is the significance of the *ascB* gene of *asc* operon which was previously considered less significant for growth on cellobiose. Here, we show through a combination of conventional genetics, adaptive evolution and targeted genome engineering that the *ascB* gene could serve as one of the most efficient β-glucosidases or even a stand-alone β-glucosidase for cellobiose metabolism in *E. coli*. In addition, we show that this gene *ascB* is being controlled by another putative promoter within the operon apart from the cryptic promoter of the *asc* operon thus opening new directions on the evolution and regulation of these operons. A combination of recombinant DNA technology and high-throughput screening process revealed that a combination of these cryptic operons could help in extending the substrate range of *E. coli* to metabolize several glucosidases including cellobiose, salicin, arbutin, gentiobiose, raffinose, and amygdalin.

It is not just sufficient to enhance the cellobiose metabolic rate in order to make a strain more proficient for

consolidated bioprocessing; it is also necessary that cellobiose is metabolized as efficiently as glucose. For instance, glucose is metabolized in a respiro-fermentative mode resulting in the secretion of large amount of acetate due to overflow metabolism whereas cellobiose is metabolized in a respirative manner secreting acetate only during the stationary phase. Such differences urge the need to understand and rewire the central carbon metabolism for an efficient cellobiose metabolism. Here we show that rewiring the flux through the TCA cycle could help in enhancing the cellobiose metabolism in *E. coli*. We demonstrate here that the transcription factor, YebK helps in enhancing the cellobiose metabolic activity through modulations in the central carbon metabolism. YebK recognizes the two major central carbon intermediates, 2-keto-3-deoxy-phosphogluconate (KDPG) and alpha-ketoglutarate (AKG), and co-ordinates the regulations of the TCA cycle of *E. coli*. Such regulation is particularly dominant during the down-shift of cells from rich nutrient source to a minimal nutrient source.

Finally, we also demonstrate the application of the strain ESS, engineered with efficient cellobiose metabolism for co-metabolism of multiple carbon sources or in consolidated bioprocessing for growth with cellulose as a sole carbon source. Thus, the strain ESS would serve as a potential host for consolidated bio-processing or in simultaneous saccharification and fermentation process. The strain ESS could also serve as a molecular bag to decipher challenging queries related to the evolution of the cryptic genes of *E. coli*. Finally, through this study we also show that the putative transcription factor binding sites or transcription start sites (TSS) reported within the coding/intergenic regions identified through ChIP-sequencing or deep RNA sequencing technologies could serve as a potential target for metabolic engineering and strain optimization.

Table of Contents

Chapter 3. Transcriptome-based characterization of strain, ESS

(Transcriptome-driven synthetic re-modelling of *E. coli* to enhance cellobiose utilization)

Chapter 4. Genotypic and phenotypic characterization of strain ESS

(Novel functions and regulations of cryptic cellobiose operons in *E. coli*)

Chapter 7. Industrial applications of strain ESS

Application II: Imposing cellulolytic ability into *E. coli*

(Engineering *E. coli* as an efficient consolidated bio-procesor)

List of Figures

List of Tables

Chapter 1

Cryptic cellobiose operons of *E. coli*

(Regulations, crypticity and potential benefits)

1.1. *Escherichia coli*- An efficiently characterized model host

Escherichia coli remains to be one of the best understood organisms till this age and continue to play a leading role in microbial research and in industrial bio-production process. Several interesting questions on metabolism, gene regulation and evolution are addressed using this simple microbe, *E. coli*. A vast majority of genome-wide analysis resources are available for *E. coli* which includes the collection of single gene deletion strains (Keio collection[1]), individual gene over-expressing plasmid library (ASKA library[2]), and genome-wide promoter-GFP library (transcriptional reporter plasmid library[3]). In addition, databases like Ecocyc[4] serves as a repository for the collection of up-to-date information of microarray or other high-throughput data[5] related to *E. coli* and maintains an updated literature on all genes present in *E. coli*. Despite vast number of researches, one-third of *E. coli* genes remain uncharacterized or orphan and new functions are continuously being reported. Examples include the recently discovered sulfo-glycolytic pathway[6], and thermal robustness of *E. coli* mediated by the temperature responsive riboswitches[7]. High-throughput screening techniques including phenotype arrays and metabolite profiling[5, 8] are used to decipher the function of unknown proteins in *E. coli*. In this study, we have employed a multi-omic approach to decipher the regulatory logics of the cryptic cellobiose operons of *E. coli*.

1.2.Cryptic cellobiose metabolic operons in *E. coli*

There are at least 4 operons (including the *chb* operon, *asc* operon, *bgl* operon and *bgc* operon) for the metabolism of β-glucoside sugars like arbutin, salicin, and cellobiose in several members of *E. coli* (Fig 1.1). These operons remain silent in *E. coli* and needs a stringent selection pressure in the presence of the β-glucoside sugar for its activation through mutations in the promoters and/or the regulatory proteins[9]. Although the reason for conservation of these operons in *E. coli* lineage is not known the cryptic nature of these operons are attributed as a way to exclude the metabolism of toxic β-glucoside sugars found in nature[10]. Understanding the regulations of these operons would help in answering several challenging questions related to the evolution of gene regulatory system and conservation of essential and/or cryptic genes in *E. coli*.

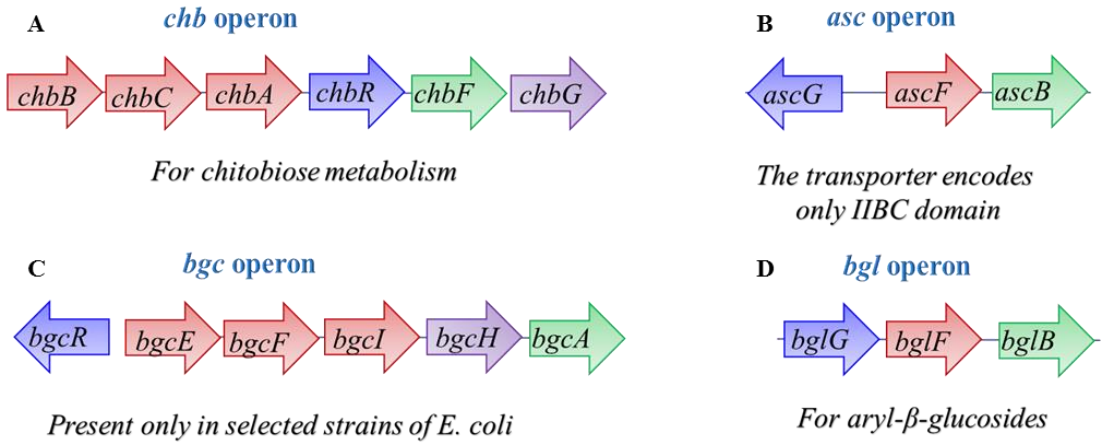


Fig 1.1 Schematic representation of the four different cryptic β-glucoside metabolic operons present in *E. coli*. The genes enclosed in red arrows encode the transporter; the genes enclosed in green arrows encode the β-glucosidase enzyme; the genes enclosed in blue arrows encode the regulatory proteins of the respective operon; the genes enclosed in purple arrows encode the protein of unknown function. (A) The native substrate for *chb* operon gene product is chitobiose but upon mutation can metabolize cellobiose, salicin, and arbutin. (B) Mutations in the *asc* operon allow growth on salicin and arbutin but not on cellobiose. (C) Mutations in the *bgc* operon allows growth on cellobiose at low temperature. (D) Mutations in the *bgl* operon allows growth on aromatic-β-glucosides: salicin and arbutin. The native substrates of *bgc*, *asc*, and *bgl* operons are not known.

The *bgl* operon was more specific for the metabolism aromatic- β-glucosides: salicin and arbutin. Different point mutations in the *bglR*, *gyrA*, *gyrB*, and *hns* genes are shown to activate the *bgl* operon or enhance the expression of *bgl* operon. In addition, the *bgl* operon was shown to be naturally expressed *in vivo* in mouse liver[2]. The activated *bgl* operon confers a growth advantage in stationary phase in strains carrying *rpoS* mutation[3]. The *bgc* operon is prevalent only among the extra intestinal pathogenic strains of *E. coli* and related Enterobacteria and few gram-positive bacteria [11]. The *bgc* operon is induced at low temperature in the presence of cellobiose or other β-glucoside sugars. However, *bgc* operon is not present in the normal laboratory strains like *E. coli* K12 and phylogenetic analysis indicates that the operon is prevalent in the ancestral Enterobacterium followed by multiple losses during species diversification [11].

The *chb* operon is not a naturally cryptic operon and is native to the metabolism of a similar disaccharide, chitobiose [12]. However, prolonged incubation of *E. coli* on cellobiose minimal medium (for approximately 14

days) leads to the isolation of spontaneous mutants that grows on cellobiose efficiently. The *chb* operon is regulated by three different transcription factors: NagC, ChbR and CAP[13]. The mechanism of mutational activation of *chb* operon for cellobiose metabolism was studied in detail and two important genetic events are reported to be essential for the activation of *chb* operon[9]. First is the insertional inactivation of NagC binding site in the *chb* promoter or loss of function of NagC protein. Deletion of *nagC* alone was sufficient to induce the *chb* promoter in response to cellobiose[13] though it could not support growth on cellobiose. The β-glucosidase, ChbF, encoded from the *chb* operon was reported to have a broad substrate range capable of hydrolyzing phosphorylated form of arbutin, salicin, gentiobiose, and cellobiose[14].

The *asc* operon allows growth of *E. coli* on salicin or arbutin upon inactivation of the repressor protein AscG, but do not allow growth on cellobiose [15]. In fact it allows growth on cellobiose minimal plates but not on liquid culture of cellobiose minimal medium. The inability of *asc* operon to confer growth on cellobiose is possibly because it encodes for an incomplete PTS transporter that lacks the A subunit [16]. Based on sequence similarity, the *asc* operon is also reported to have been originated from the duplication of *bgl* operon of *E. coli* [15]. The *asc* operon is considered as a semi-cryptic operon and the *ascG* is also shown to be regulating propionate metabolism in *E. coli*[17].

Of the several β-glucosides, the metabolism of plant derived β-glucoside, cellobiose is more important for the establishment of an industrially valuable host [18]. Engineering efficient cellobiose metabolism using the native genes of *E. coli* would help in reducing the number of heterologous cellulolytic system that would be needed to make *E. coli* an efficient catalyst for lignocellulosic biomass. This would help in achieving a balance between the heterologous system and the innate metabolic potential[19]. By far, the *chb* operon was considered to be the major contributor for cellobiose metabolism in *E. coli*. With the advent of synthetic biology vast majority of regulated and constitutive promoters are available to engineer the genes directly on the chromosome[20].

Table 1.1. Comparison of the properties of two different phospho-β-glucosidases present in *E. coli* K12.

Properties	ChbF	AscB
Protein Family (CaZY group)	Family 4	Family 1
Co-factor	NAD+	-
Native substrate	Chitobiose	-
Cryptic substrates	Cellobiose, arbutin, salicin, gentiobiose	Cellobiose, salicin, arbutin
K_m for PNPG	0.44 mM	0.21 mM*

* Based on a similar protein from Pectobacterium (82 % similar)[4]

1.3. Industrial benefits of engineering cellobiose metabolism in *E. coli*

Human society relies extensively on machines to perform even simple tasks thus making fuels an indispensable requirement for sustainability. The rapid increase in global energy demand and environmental impacts combined with decrease in the availability of the current energy sources argues to forgo fossil fuels and search for alternative, cleaner, renewable and sustainable energy sources such as solar energy, hydroelectric energy, wind energy, and bioenergy[21]. In recent days, the production of bioenergy from lignocellulosic biomass has a significant global interest[22]. Starch and sucrose have been used as the sugar source for biofuel production till date[23]. However, the use of starch or sucrose is disadvantageous in that special fuel crops should be reared. Demand for dedicated fuel crops make the process economically challenging and also may lead to food versus fuel conflict[23]. This disadvantage could be overcome by exploiting the hidden sugar treasure: **lignocellulose**, the most abundant and renewable carbon deposit on earth[22].

Lignocellulose is a complex and recalcitrant polysaccharide that goes to make up the plant cell wall and is composed of three components: celluloses, hemicelluloses and lignin. Cellulose is a homopolymer of cellobiose whereas hemicellulose is a heteropolymer of sugars like mannose, galactose, xylose, arabinose, rhamnose and fucose. While both cellulose and hemicellulose are made up of fermentable sugars, lignin consist of aromatic polymers especially, toxic phenolic compounds[24]. Lignocellulose is one of the most recalcitrant polymers on the Earth. Hence, the efficiency of lignocellulosic bio refinery is largely dependent on the microbial catalyst employed in this process. Developing a microbial host with all peculiar traits needed for the efficient conversion of lignocellulosic substrate to biofuels is far more complex than it was previously expected (Fig 1.2). Since all the traits needed for conversion of lignocellulose to biofuels cannot be found within a single microbe (of the microbes

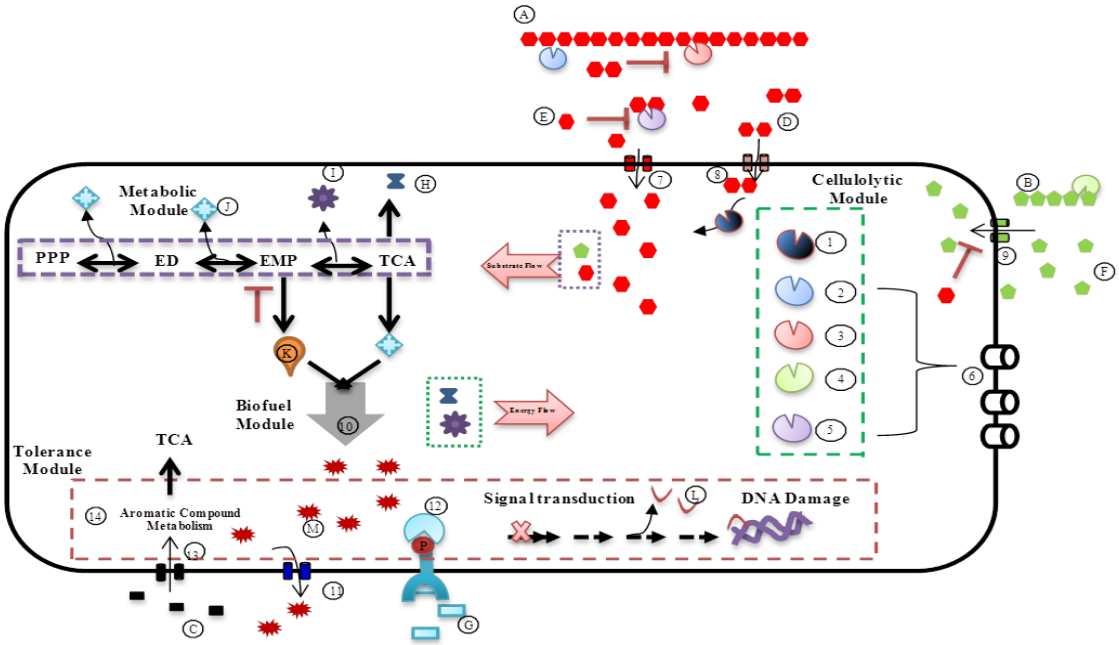
reported so far), integration and co-ordination of several heterologous modules within a single host remains to be a biggest challenge to synthetic biologists and metabolic engineers.

One approach in developing a microbial host that could degrade lignocellulosic biomass efficiently is to mimic the native cellulolytic microbes. In accordance with that, the native cellulolytic microbes obtain all energetic benefits for growth on cellulose through cellodextrin/cellobiose metabolism[25]. Synergistic action of at least 3 enzymes (endoglucanase, exoglucanase, and β-glucosidase) is required to produce glucose from cellulose. Cellobiose-metabolizing strains would help eliminate the need for secretion of one (β-glucosidase) of the 3 enzymes described above [26]. Cellobiose is the major source of carbon for soil bacterial community[27]. In addition, rapid uptake of cellobiose and reduced extracellular glucose level will help alleviate the glucose inhibition of cellulolytic enzymes[28]. For example, most of the cellulase enzymes suffer from feedback inhibition by cellobiose and glucose, thus decreasing the overall productivity significantly [29]. In addition, engineered cellobiose metabolism could help in overcoming carbon catabolite repression between the metabolism of glucose and xylose and allows cells to co-metabolize cellobiose with xylose[30]. Hence, engineering efficient cellobiose metabolism is an important prerequisite to develop an industrial host for biofuel production from lignocellulosic substrate.

Several industrial microorganisms, including *Zymomonas mobilis*, *Klebsiella oxytoca*, *Escherichia coli*, and *Saccharomyces cerevisiae*, have been proposed to be suitable candidates for Consolidated Bioprocessing, a single step process that allows the conversion of lignocellulosic substrate to biofuels without any added enzyme but with a microbial catalyst[31]. Among these microbes, *E. coli* is of particular interest because of its natural ability to utilize both hexose and pentose sugars derived from plant biomass. It is a well-known industrial bacterium with a high growth rate and metabolic activity, it has been engineered extensively to produce several advanced biofuels other than ethanol, and abundant molecular biological tools are available for its genetic manipulation [32]. In addition, *E. coli* contains cryptic genes for cellobiose utilization [33]. Although wild-type *E. coli* cannot utilize cellobiose, it can acquire the cellobiose-utilizing (Cel+) phenotype after a period of starvation in cellobiose-containing minimal medium[9, 33-34]. The Cel+ phenotype has been found to be linked to mutations in two cryptic operons, *chb* or *asc* [16].

Cellobiose metabolism was engineered in *Escherichia coli* using heterologous genes from *Klebsiella* sp. [35]. Similarly, phosphorolytic cellobiose metabolism was achieved using *Saccharophagus* β-glucosidase and *E. coli*

native lactose transporter [36]. All of these engineered *E. coli* strains grew on cellobiose (either with cellobiose as the sole carbon source or supplemented with yeast extract), albeit slowly. In this study, we aimed at exploiting the native cryptic operons for cellobiose metabolism for two reasons: (1) to reduce the number of heterologous cellulolytic system that should be engineered into the microbial host and (2) strike a balance with the native microbial potential [19].



Traits	Cellulolytic microbes	Solventogenic microbes	Reaction steps	Limiting metabolites
Cellulase production	+	–	1-5	A, B, I,H
Protein Secretion System	+	–	6	I
Advanced biofuel production	–	+	10	K, J
Lignin tolerance	+	–	12-14	C, G, L
Solvent tolerance	–	+	11	L, M
Elimination of CCR	+/-	+/-	7-9	D, E, F

Fig. 1.2. Scheme for the minimal requirements of an ideal microbe for consolidated bioprocessing of lignocellulosic biomass. An ideal CBP microbe should be capable of expressing and secreting several glycoside hydrolase enzymes, and processively metabolize soluble sugars like cellobiose, glucose and xylose to biofuels. In addition, the ideal microbes should have tolerance against lignin-derived compounds and the biofuels. Substrates and Products: A, Cellulose; B, Hemicellulose; C, Lignin dervied products; D, Cellobiose; E, Glucose; F, Xylose; G, sugar degradion products; H, Amino acid; I, ATP; J, NADH/NADPH; K, Acetyl coA; L, Reactive oxygen

species; M, Biofuels. Enzymes and Metabolic pathways: 1, Intracellular β-glucosidase; 2, Exoglucanase; 3, Endoglucanase; 4, Hemicellulase; 5, Extracellular β-glucosidase; 6, Protein secretion system; 7, Glucose transporter; 8, Cellobiose transporter; 9, Xylose transporter; 10, Biofuel producing pathway; 11, Solvent export pump; 12, engineered mutant transcription factor; 13, phenolics transporter; 14, Aromatic compound degradation pathway.

1.4. Objective

The main objective of this study is to reactivate the cryptic cellobiose metabolic ability of *E. coli* using well-established promoter elements and understand/rewire the regulatory networks governing cellobiose metabolism in *E. coli*. This study would help in answering several evolutionary and molecular mysteries behind the cryptic cellobiose metabolic operons in *E. coli* and also help in developing platform strains for lignocellulose-based biorefineries. Through this study, we also show that the putative internal Transcription Start Sites (TSS) or the transcription factor binding sites within the coding regions of operons (that are increasingly been reported through the high-throughput screening techniques) could serve as one of the potential targets for metabolic engineering or strain optimization for efficient growth on cellobiose.

The two major objective of this study is: (1) Understanding the regulatory network governing the cryptic cellobiose metabolism (2) Developing a platform strain for consolidated bioprocessing. A combination of recombinant DNA technology (through promoter engineering), adaptive evolution and high-throughput screening helps in addressing the function of cryptic genes of *E. coli* in greater depths than with the conventional genetics.

Chapter 2

Design and construction of strain ESS

(Engineering *E. coli* for efficient cellobiose metabolism)

2.1 Abstract

Escherichia coli normally cannot utilize the β-glucoside sugar, cellobiose as a carbon and energy source unless a stringent selection pressure for survival is present. The cellobiose-metabolizing phenotype can be conferred to wild type *E. coli* through mutations in two cryptic operons: *chb* and *asc*. In this study, the cellobiose-utilization phenotype was conferred to *E. coli* by replacing the cryptic promoters of these endogenous operons with a constitutive promoter thus removing all regulatory sequences that are reported to keep the *chb* and the *asc* operons cryptic. Evolutionary adaptation of the engineered strain OSS (Original Synthetic Strain) by repeated subculture in cellobiose-containing minimal medium led to an increase in the rate of cellobiose uptake and cell growth on cellobiose resulting in a strain, ESS (Evolved Synthetic Strain). The growth rate of strain ESS on cellobiose minimal medium was more similar to the growth rate on glucose minimal medium. The strains, OSS and ESS, are different from previously obtained cellobiose-metabolizing strains of *E. coli* in that the later strains are based on stress-induced mutations in the *chb*/*asc* operons.

2.2. Introduction

As described in Chapter 1, there are at least two cryptic operons for cellobiose metabolism in *E. coli*. Despite being a non-cellulolytic microorganism, *E. coli* possesses a set of cryptic genes involved in cellobiose metabolism [9]. These encoded genes are indicative of cellulolytic ability in *E. coli*, which presumably has been lost during evolution. Several spontaneous Cel+ mutants of *E. coli* have been isolated previously, which suggests that *E. coli* has not lost all genes needed for cellulose utilization. The main reason for the crypticity of these operons was attributed to the promoters of these operons. In accordance with that the mutations in the regulatory proteins (such as *chbR*, *nagC*) or insertion of IS elements within the promoters' leads to the activation of these operons and allows an inducible growth on cellobiose. Previous studies were based on the naturally acquired mutations in the *chb* or *asc* operon. In addition, previous studies focused on characterizing the function of either of these operons (independently). Hence, the combined ability of these two operons still remains unexplored. Especially, the potential of *asc* operon was undermined with respect to cellobiose metabolizing ability because it encodes only for an incomplete transporter. We exploited these cryptic genes involved in cellobiose metabolism by rationally engineering them to confer a cellobiose-utilization phenotype in *E. coli*.

This work reports the design, construction, and optimization of Cel+ *E. coli* strains by replacing the promoters of the cryptic endogenous *chb* and *asc* operons with a synthetic constitutive promoter, and by evolutionary adaptation of these strains for rapid cellobiose metabolism. We characterized the ability of the *chb* and *asc* operons independently or when activated together. We obtained a strain, designated ESS, which exhibited a similar growth rate in minimal medium supplemented with either cellobiose or glucose, through adaptive evolution. The strain ESS would help in addressing several challenges related to evolution and crypticity of the *chb* and *asc* operon especially because unlike the previously obtained natural cel+ variants, strain ESS is not obtained based on stress-induced mutations.

2.3. Materials and methods

2.3.1. Bacterial strains and media composition

The strains used are listed in Table 2.1 *E. coli* MG1655 was used as a parental strain. Bacteria were cultured at 37℃ in Luria Bertani broth (LB) or minimal medium supplemented with cellobiose or glucose. Strains carrying temperature-sensitive plasmids were grown at 30℃. Media were supplemented with suitable antibiotics (50 µg

kanamycin/mL and 100 µg ampicillin /mL). For long-term storage, cells were maintained as 20% glycerol stocks at –80℃. M9 minimal medium, supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 4 g sugar/L, was used to characterize the cell growth rate of the modified strains. In these tests, 1% of overnight cultures grown in LB were inoculated into 50 mL of M9 medium supplemented with the test sugar in a 250-mL flask, and the cultures were incubated at 37℃ with rotation at 200 rpm. Cell growth was monitored by measuring the optical density at 600 nm (OD_{600}) with a Biochrom Libra S22 spectrophotometer. Solid minimal medium, supplemented with cellobiose, contained 1% (w/v) agar.

Table 2.1 *E. coli* strains and plasmids used in this study

Strains/ plasmids	**Description/genotype**	**Reference/ source**
Strain		
E. coli MG1655	Wild-type	[37]
OSS-C	MG1655 with CP12*chb*	This study
OSS-A	MG1655 with CP12*asc*	This study
OSS	MG1655 with CP12*chb* plus CP12*asc*	This study
ESS-C	OSS-C adapted in cellobiose for 75 days	This study
ESS	OSS adapted in cellobiose for 30 days	This study
Plasmids		
pKD46	λ-Red recombinase expression plasmid and temperature sensitive replication	[38]
pCP20	Yeast FLP recombinase gene controlled by *c*I repressor and temperature sensitive replication.	[39]
pKD13	Template plasmid for gene disruption. The resistance gene is flanked by FRT sites. *oriR6K*-gamma origin requiring the *pir*$^+$ *E. coli.*	[38]

2.3.2. Strain construction

Promoter replacement was performed using the λ-Red recombination system, as described previously [38]. The synthetic constitutive promoter (CP12) used in this study has been characterized previously and was reported to

have a β-galactosidase activity of 101 miller units in *E. coli* [40]. The synthetic promoter was generated by PCR. For promoter replacement, two overlapping fragments were amplified using Splice Overlap Extension (SOE) PCR. Fragment 1 contained the promoter, with primer homologous to the downstream region of the endogenous promoter. Fragment 2 contained the kanamycin cassette from pKD13, with an overhang that is homologous to the upstream sequence of the endogenous promoter. Cells carrying the λ-Red system, under control of the *araBAD* promoter (pKD46), were induced with 10 mM arabinose, made electrocompetent, and then transformed with the PCR products. Transformant colonies carrying the desired modification were directly selected on LB agar plates supplemented with kanamycin. Genomic DNA was isolated from the transformants, and the target region was PCR-amplified and sequenced to confirm site-specific insertion. The primers used for strain construction are listed in Table 2.2.

The cryptic promoter of the *chb* operon in *E. coli* MG1655 was replaced with the synthetic constitutive promoter (CP12) described above, and the strain so generated was designated OSS-C. In strain OSS-A, the promoter of the *asc* operon was replaced with CP12. Strain OSS was constructed by replacing the promoters of both *asc* and *chb* operons with the constitutive promoter CP12.

Table 2.2 Primers used in the study

Name	Primer sequence
SOEing CP12 promoter to Kanamycin cassette	
	5′- CATAGCTGTTTCCTGTGTGAACAGTACTCAGGTATTATATCATTTTG -3′
	5′- TCAGGTATTATATCATTTTGGCCGACTAGTGTCAAGAATAAACTTG -3′
	5′- TAGTGTCAAGAATAAACTTGTATATGATTCCGGGGATCCGTCGACC -3′
Promoter exchange	
CP12CHB	5′- AAAGCTTGAATAACAACGGAAACCGGCCATTGCGCCGGTTTTTTTTCTGAGTTGTGTAG GCTGGAGCTGCTTCG -3′
	5′- CACTCCGCAAGCCTTAACCTGCTTCCATGCTCTGGGTAACTTGCGAAACATAGCTGTTT CCTGTGTGAACA -3′
CP12ASC	5′- CGATTGTCCTTGCACAATCGGCGGGAAAAATATTCAGGTGACCGGTTTCAGTGTAGGCT GGAGCTGCTTCG -3′
	5′- ATCAACGCCGCCCAGTGCCGCTATCACCGAGCGTGCCAGCGCCGCATAATTTTTGGCCA TAGCTGTTTCCTGTGTGAACA -3′

Primers for sequencing	
CP12CHB	5′- GGTACTTGCCAGACCTACATCCTGG -3′
	5′- TGCAACCGCAGCCTTAAGC -3′
CP12ASC	5′- ATGATGTCATCGATCTCATCCACGC -3′
	5′- TGACGACTTCCTGAAAGGCTTGTGA -3′

2.3.3. Enzyme assays

β-glucosidase assays were performed by measuring the release of *p*-nitrophenol, as described previously [41]. In brief, 2 mL of cultures grown overnight in LB at 37℃ were lysed and suspended in 200 µl of 50 mM sodium phosphate buffer (pH 7.0). Then, 100 µl of this crude cell extract was incubated with 400 µl of 10 mM *p*-nitrophenyl-β-D-glucopyranoside (PNPG) at 37℃ for 2 hours. 1 ml of 1 M Na_2CO_3 was added to stop the reaction, and the absorbance of the liberated *p*-nitrophenol was measured at 410 nm. One unit of enzyme activity is defined as the amount of enzyme that catalyzes the conversion of 1 µmol of PNPG per min. Absorbance was measured with a Biochrom Libra S22 spectrophotometer.

2.3.4. Adaptive evolution

The strains OSS-C and OSS were adapted in minimal medium supplemented with cellobiose. Cells were transferred to fresh medium every time the culture reached an OD_{600} of 1.0. The effect of the adaptation on cell growth was analyzed every 30 days. Adaptation of the slow-growing strain OSS-C was performed for 75 days, and adaptation of OSS for 30 days.

2.3.5. HPLC analysis

The residual sugar in the culture medium during the bacterial growth period was measured regularly using a Shimadzu HPLC station, equipped with an HPX-87P column (Bio-Rad) and a refractive index detector (Shimadzu). HPLC-grade pure water, at a flow rate of 0.6 mL/min, was used as the mobile phase. The oven temperature was maintained at 80℃. A standard curve was determined with varying concentrations of glucose and cellobiose.

2.4. Results and discussion

2.4.1. Activation of the cryptic genes for β-glucoside metabolism in *E. coli*

The cryptic promoters of the *chb* and the *asc* operons were successfully replaced with a synthetic constitutive promoter as described in Materials and methods. Functionality of the synthetic promoters for both operons were verified in strains OSS-C, OSS-A, and OSS by β-glucosidase assay (Fig. 2.1). β-glucosidase activity of strain OSS-A was 3-fold lower than that of strain OSS-C (Fig. 2.1). Strain OSS exhibited higher enzyme activity that is equivalent to the sum of the enzyme activities in both strains, OSS-C and OSS-A (Fig. 2.1). There is no significant variation in the expression levels of the *chb* and the *asc* operons under different culture conditions that were tested (including cells grown on LB or M9-Gluocse), which indicated that the synthetic promoter is not affected by cellular metabolism.

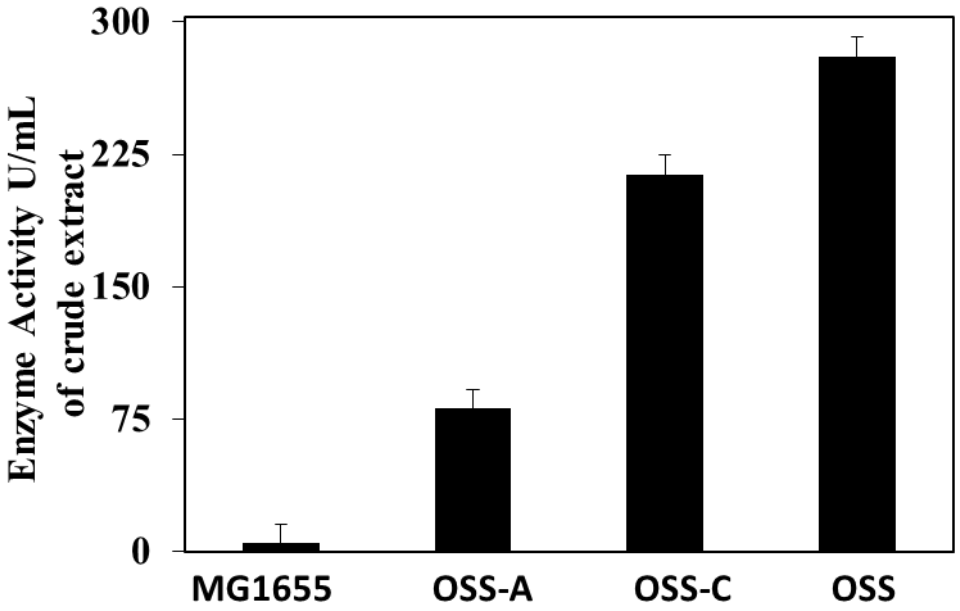


Fig. 2.1. Comparison of β-glucosidase activities among engineered cellobiose-utilizing *E. coli* strains (OSS-A, OSS-C, and OSS) and wild-type MG1655. Enzyme activity is expressed in units per mL of the crude cell extract. Error bars indicate the standard deviation of triplicate determinations.

2.4.2. Cell growth of the engineered cellobiose-utilizing *E. coli*

In glucose-containing minimal medium, the engineered strains OSS-C, OSS-A, OSS and the wild-type strain

MG1655 exhibited a similar growth rate, indicating that the constitutive expression of the *chb* and the *asc* operons may not cause significant metabolic burden to the engineered strains. Culture growth rates of the engineered strains (OSS-C, OSS-A, and OSS) in cellobiose-containing minimal medium were then analyzed. Mere replacement of the promoter of the *chb* operon with a synthetic constitutive promoter was sufficient to confer a Cel+ phenotype. This observation is consistent with the previous results showing that the cryptic nature of the *chb* operon is linked to its promoter [9]. The rate of cellobiose utilization in OSS is two-fold higher than in OSS-C indicating that the *asc* operon, as previously suspected, might not be a less specific operon for cellobiose utilization.

Although the strains OSS-C and OSS were able to utilize cellobiose as a sole carbon source, the growth rate was not as high as that obtained with MG1655 in glucose-containing minimal medium (Fig. 2.2). On the other hand, strain OSS-A could not grow in cellobiose-containing minimal medium. This phenotype of strain OSS-A is consistent with the previous observation made in the spontaneous Cel+ *asc* mutants of *E. coli* lacking both the *bgl* and *chb* operons [42]. These results indicate that the lack of growth of strain OSS-A on cellobiose-containing minimal liquid medium might be because of the transporter (AscF) that lacks sufficient subunits to form a complete PTS for cellobiose transport. In theory, exchanging the native promoter of the *asc* operon with a constitutive promoter in strain OSS-C should provide an additional copy of a glucosidase and a transporter to the strain, and thus, improve growth on cellobiose; this hypothesis was confirmed in strain OSS, which exhibits a higher growth rate in the presence of cellobiose than OSS-C (Fig. 2.2).

Use of a synthetic constitutive promoter for engineering cellobiose metabolism may have many uses in industrial applications, as it does not need an inducer. To the best of our knowledge, this is the first report of engineering the endogenous cryptic genes for the utilization of cellobiose in *E. coli*. Cellobiose is a natural substrate for *K. oxytoca*. The *casAB* operon in *K. oxytoca* has been expressed in *E. coli* KO11; however, this was not sufficient to support growth of the recombinant *E. coli* in cellobiose-containing medium[43].

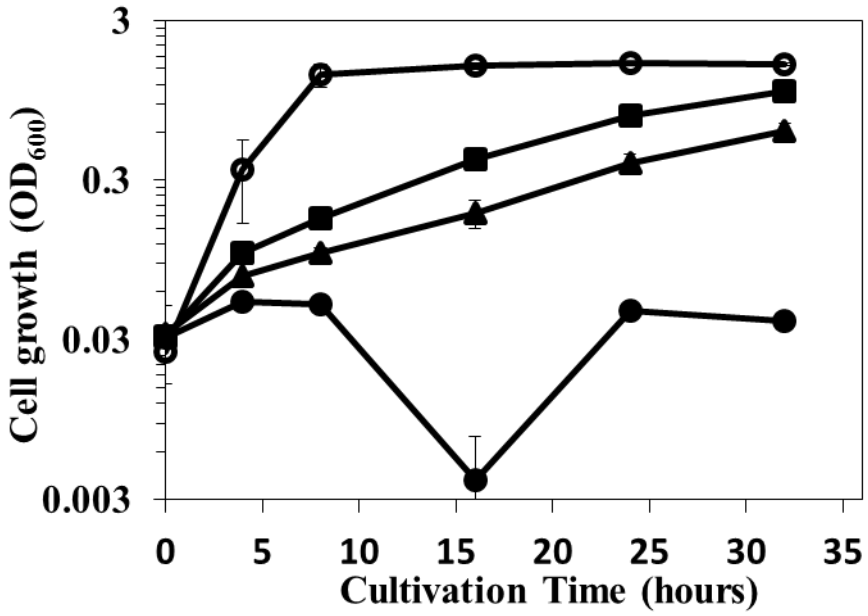


Fig. 2.2. Comparison of cell growth between wild-type *E. coli* strain MG1655 (filled circles), and the engineered *E. coli* strains OSS-C (filled triangles) and OSS (filled square) in 4 g cellobiose/L of minimal medium; and MG1655 (open circle) in 4 g glucose/L of minimal medium. Error bars indicate the standard deviation of experiments performed in triplicates.

2.4.3. Adaptive evolution in the presence of cellobiose

Though cellobiose is a disaccharide of glucose, the growth rates of strains OSS-C and OSS in cellobiose-containing minimal medium were lower than that of the parental strain MG1655 in glucose-containing minimal medium (Fig. 2.2). *E. coli* MG1655 growing on higher concentration of glucose is known to suffer from growth inhibition due to the secretion of large quantity of acetate [44]. Because cellobiose is a disaccharide of glucose, 1 molecule of cellobiose is theoretically equivalent to 2 molecules of glucose. We tested if the concentration of cellobiose (4 g/L) used in this study causes a substrate-mediated growth inhibition in strains OSS-C and OSS. The growth rates of OSS-C and OSS increase dramatically with increase in cellobiose concentration from 1 to 6 g/L, indicating that the lower growth rate of the OSS-C in cellobiose is not due to a growth-inhibiting concentration of cellobiose (Fig. 2.3a and 2.3b).

Despite providing an additional copy of the transporter and a glucosidase, the cellobiose utilization by strain OSS was still slower than glucose utilization by MG1655. This might be because *E. coli* has lost the ability

to metabolize cellobiose during evolution. In order to regain this ability, the strains OSS and OSS-C were adapted on cellobiose-containing minimal medium for 30 and 75 days, respectively, in order to achieve efficient growth with cellobiose as a sole carbon source (Fig. 2.3c); those adapted strains are referred to as ESS and ESS-C. There was a profound increase in growth rates of both strains after 30 days of adaptation. Though the growth rate of strain ESS-C in cellobiose surpassed the growth rate of OSS, it was still lower than the growth rate obtained in glucose; therefore, ESS-C was adapted further (Fig. 2.3a). However, there was no significant difference in the growth rates between 30, 60, and 75 days of adaptation of strain OSS-C, which indicated that it had attained a stable rate of growth in cellobiose-containing medium. The growth rate of OSS in cellobiose was equivalent to its growth rate in glucose after 30 days of adaptation (Fig 2.5). Hence, strain ESS was used for further studies.

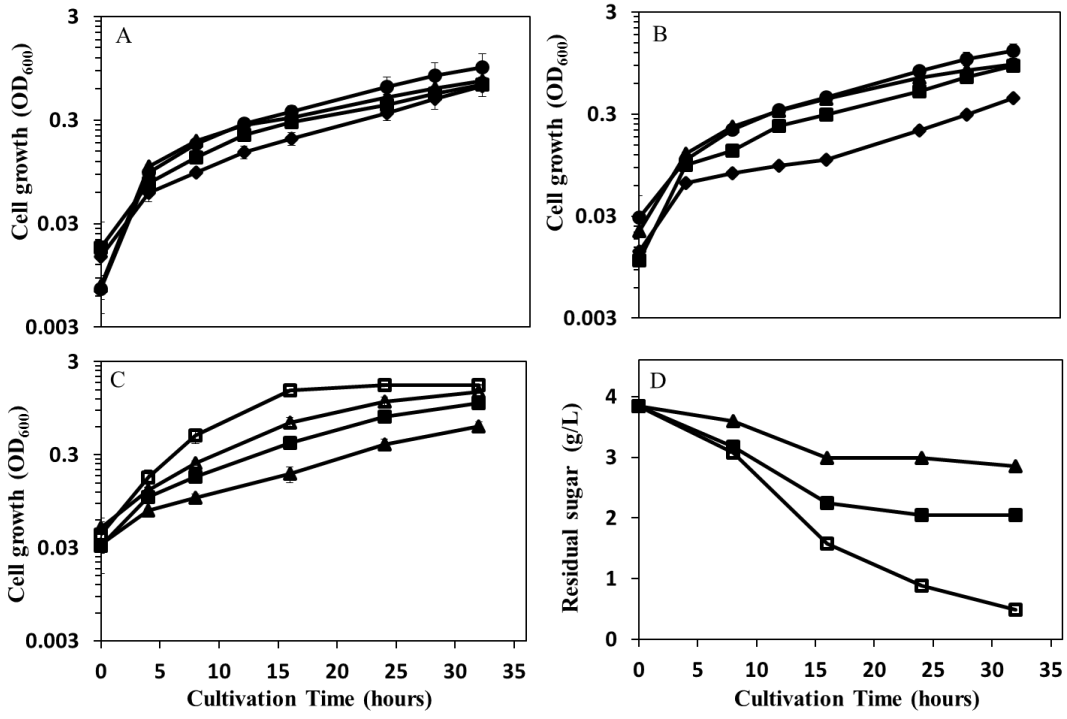


Fig. 2.3. Comparison of growth rates of the engineered strains OSS-C (**A**) and OSS (**B**) in minimal media supplemented with varying concentrations of cellobiose. 1 g/L, *closed diamond*; 2 g/L, *closed square*; 4 g/L, *closed triangle*; 6 g/L, *closed circle*. Comparison of cell growth between the engineered strains OSS-C (*filled triangle*) and OSS (*filled square*), and the adapted strains ESS-C (*open triangle*) and ESS (*open square*), in cellobiose-containing minimal medium (**C**). Error bars indicate the standard deviation of experiments performed in triplicates. Comparison of the cellobiose utilization rates between the engineered strains OSS-C (*filled triangle*)

and OSS (*filled square*) and the adapted strain ESS (*open square*) (**D**).

2.4.4. Rate of cellobiose utilization

We compared the rate of cellobiose utilization between the engineered strains, OSS-C and OSS, and the adapted strain ESS. The initial rate of cellobiose utilization was the same in all the strains. The rate of utilization of cellobiose increased with time in the adapted strain ESS (Fig. 2.3d). After 32 hours, strains OSS-C, OSS, and ESS utilized 0.8 g/L, 1.8 g/L, and 3.3 g/L of cellobiose, respectively. Thus, there was a 2-fold increase in the amount of cellobiose utilized by the adapted strain. There was no significant increase in the β-glucosidase activity between strains OSS and ESS (Fig. 2.4).

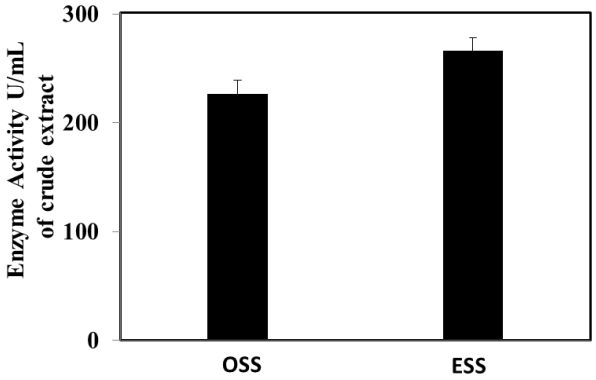


Fig. 2.4. Comparison of β-glucosidase activities between the engineered strain OSS and the adapted strain ESS. Error bars indicate the standard deviation of experiments performed in triplicates.

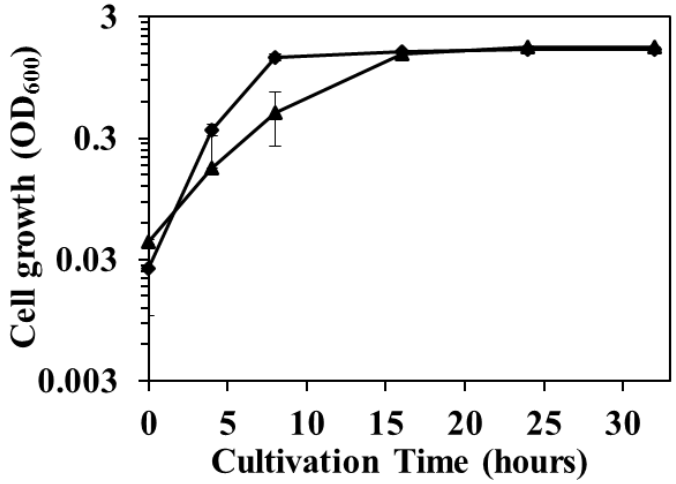


Fig. 2.5. Comparison of growth of strain ESS on minimal medium containing glucose (solid square) or cellobiose

(solid triangle). Error bars indicate the standard deviation of experiments performed in triplicates.

2.5. Conclusions

We have genetically engineered and adapted *E. coli* for rapid cellobiose utilization. The time required to obtain a spontaneous Cel+ mutant by conventional selection is too long, and the mutants obtained do not sufficiently express the *chb* operon to support rapid cell growth on cellobiose. In addition, the genotypes of spontaneous Cel+ mutants have not been fully characterized. These reasons could explain why no further progress has been made in cellulosic fuel production with spontaneous Cel+ mutants. Also, *E. coli* strains reported earlier, which carry heterologous genes for cellobiose metabolism, have been shown to be incapable of rapid cellobiose fermentation [43b]. However, the engineered strains described in the present studies (OSS-C and OSS) have more stable genotypes, and exhibit efficient cellobiose-utilization phenotypes. The genotypes of the engineered Cel+ OSS strain can be fine-tuned further for the production of desired products. Heterologous cellobiose metabolism has also been reported in *S. cerevisiae* by expressing the β-glucosidase gene from *Neurospora crassa* [30], and the ethanol yield obtained from cellobiose metabolized by this engineered yeast strain was found to be slightly less than that from glucose. Cellobiose metabolism also relieved this recombinant yeast strain from catabolite repression by providing a spatial separation of glucose and other sugars. Co-consumption of xylose and cellobiose was also demonstrated with this strain [30]. Hence, this recombinant Cel+ strain might be a promising platform host for developing a CBP strain. Cellobiose is the major soluble by-product in the enzymatic saccharification of cellulose [45].

Adaptation of the strain OSS for rapid cellobiose utilization showed that *E. coli* can be engineered, without the introduction of heterologous genes, for cellobiose utilization that is as efficient as glucose utilization. Understanding the rate-limiting step in cellobiose metabolism would help in designing and constructing a tailor-made strain for optimal cellobiose utilization. Such strains could help overcome cellobiose inhibition of endo- and exoglucanases in the SSF process. They might also help in the elimination of glucose effects on hemicellulosic sugar utilization, because intracellular cellobiose metabolism provides a spatial separation of glucose and other sugars [30]. The active utilization of cellobiose as a single carbon source by the adapted strain ESS should be of great interest for lignocellulosic fuel production.

Chapter 3

Transcriptome-based characterization of strain, ESS

(Transcriptome-driven synthetic re-modeling of *Escherichia coli* to enhance cellobiose utilization)

3.1. Abstract

In this study, gene expression patterns in rapid (ESS) and slow (OSS) cellobiose-metabolizing strains were analyzed using DNA microarray. Genes involved in nucleotide biosynthesis, ATP generation, and flagellar assembly were up-regulated following adaptation, indicating the possibility of active metabolism and fast growth. On the other hand, genes involved in chemotaxis toward glucose and response to starvation and other stresses were down-regulated. No significant changes were observed in the expression of genes regulated by cyclic AMP receptor protein (CRP), the global regulator of carbon metabolism. Overexpression of *ascB*, which is located in the minor cryptic operon, seems to play a vital role in enhancing cellobiose metabolism. Changes in *ascB* transcription were confirmed by semi-quantitative RT-PCR. The importance of *ascB* for cell growth on cellobiose was also validated by overexpression and knockout studies. Our study provides new insights into the role of the *chb* and *asc* operons in cellobiose metabolism.

3.2. Introduction

As described above, cellobiose metabolism was achieved by exploiting native *chb* and *asc* operons. Promoters of these cryptic operons were replaced with a synthetic constitutive promoter in wild-type *E. coli* generating a slow cellobiose-metabolizing strain (OSS) [26]. Laboratory-scale adaptive evolution of the synthetic strain (OSS) on cellobiose minimal medium for 30 days gave rise to a robust cellobiose-metabolizing strain (ESS) [26]. The main disadvantage of using adaptive evolution for strain improvement or for further molecular characterization is the possible accumulation of mutations in genes that are not associated with the phenotype of interest [46]. Understanding the molecular, regulatory, and genotypic changes during adaptive evolution might pave the way for the rational design of synthetic strains [47].

Systematic analysis of global gene expression profiles may be a powerful tool in predicting the response of bacteria to environmental stimuli [48]. Microarray technology is thus a superior method to measure the global pattern of differentially expressed genes across different strains [49]. Transcriptional regulation is the first line of defense enforced by bacteria for the efficient utilization of resources [50]. A comparison of transcriptomic polymorphisms between strains with different uptake rates of a particular carbon source would help us understand the rate-limiting step and changes in metabolic state. It is a well-established fact that adaptive evolution helps strengthen pre-existing pathways rather than substituting new enzymes for a better phenotype. Hence, changes in mRNA abundance could indicate the key regulatory events that increase the fitness of evolved strains.

In this study, we compared global gene expression patterns in OSS and ESS in order to understand and determine the rate-limiting steps of cellobiose metabolism. Transcriptome profiling of ESS and OSS provided new insights into the crypticity of the *chb* and *asc* operons and introduced key ideas for reverse engineering of a tailor-made strain of *E. coli* capable of rapid cellobiose metabolism. Differences in flux through the central metabolic pathway (CMP) of strains metabolizing cellobiose and glucose, respectively, were also predicted on the basis of relative mRNA abundance. These data could also reveal the possible routes of metabolic flux that could favor co-metabolism of hexose and pentose sugars.

3.3. Materials and methods

3.3.1. Bacterial strains and culture conditions

The strains used are listed in Table 3.1. *E. coli* MG1655 was used as a parental strain. Bacteria were cultured at 37℃ in Luria–Bertani broth (LB) or minimal medium supplemented with 4 g/L cellobiose. Strains carrying temperature-sensitive plasmids were grown at 30℃. Media were supplemented with suitable antibiotics (50 μg/mL kanamycin, 30 μg/mL chloramphenicol, or 100 μg/mL ampicillin). For long-term storage, cells were maintained as 20% glycerol stocks at −80°C. Protein expression was induced by addition of tetracycline (1–5 ng/mL) or sodium propionate (20 mM). M9 minimal medium supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 4 g/L cellobiose was used to characterize growth of the modified strains. Initial experiments to optimize the inducer concentration were performed in 96-well plates, and cell growth was monitored every hour using a Tecan Infinite F200 PRO plate reader (Tecan-US, Durham, NC). Further validations were performed in 250-mL shake flasks, in which cells were inoculated (1:100) from overnight LB cultures into 50 mL M9 medium supplemented with the test sugar; the cultures were incubated at 37℃ with rotation at 200 rpm. Cell growth was monitored by measuring the optical density at 600 nm (OD_{600}) with a Biochrom Libra S22 spectrophotometer.

Table 3.1. *E. coli* strains and plasmids used in this study

Strains/plasmids	Description/genotype	Reference/source
Strain		
E. coli MG1655	Wild-type	[37]
OSS	MG1655 with CP12*chb* plus CP12*asc*	[26]
ESS	OSS adapted in cellobiose for 30 days	[26]
OSS-C	MG1655 with CP12*chb*	[26]
OSS-A	MG1655 with CP12*asc*	[26]
ESS-B	ESS with Δ*ascB*	This study
ESS-F	ESS with Δ*chbF*	This study
ESS-T	ESS with Δ*chbBAC*	This study
Plasmids		
pSIM5	λ-Red recombinase expression plasmid and temperature sensitive replication	[51]
pCP20	Yeast FLP recombinase gene controlled by *c*I repressor and temperature sensitive replication.	[38]

pKD13	Template plasmid for gene disruption. The resistance gene is flanked by FRT sites. *oriR6K*-gamma origin requiring the *pir*$^+$ *E. coli.*	[38]
pZB	Cm^r; p15A replicon, PBAD promoter, P_{zt1} promoter	[52]
pZB-G	pZB plasmid with *pgi*, *fbaA*, *gapA* and *tpiA* expressed from P_{zt1}	This study
pZB-A	pZB plasmid with *ascB* expressed from P_{zt1} and with strong RBS (AAGAAG)	This study
pZB-C	pZB plasmid with *chbF* expressed from P_{zt1} and with strong RBS (AAGAAG)	This study
pPro-*ascB*	pPro29b plasmid expressing *ascB* from P_{pro}	This study
pPro-*chbBAC*	pPro29b plasmid expressing *chbBAC* from P_{pro}	This study

3.3.2. Transcriptome analysis

For transcriptome analysis, OSS and ESS were grown on minimal cellobiose medium up to an OD_{600} of 0.7 as described previously in order to compare the central metabolic changes for strains grown on glucose, acetate or glycerol[53]. Cells were harvested from 500 mL of the culture and washed with 100 mM sodium phosphate buffer; the pellet was stored at –80℃ until analysis. Transcriptome profiling was performed using the GE microarray, 3*20K (*E. coli* str K12 substr, MG1655) platform as per the manufacturer's protocol. Three biological replicates were analyzed for each sample. Genes with signal intensities greater than 100 were considered for further analysis. Average signal intensity from triplicate readings of each ESS gene was normalized to the corresponding average signal intensity of OSS, and fold-change values were obtained. For functional analysis of the transcriptome, we applied ADGO [54] to the fold-change values. Three categories of Gene Ontology (GO) were chosen simultaneously for functional interpretation, and the Z-statistic method was used to summarize the fold-change values of each GO term. We identified 175 up-regulated and 70 down-regulated GO terms. Representative categories obtained from ADGO are summarized in Table 3.2. A list of genes involved in the CMP (as per the iAF1260 model) [55] was considered for experimental validation. Heat maps were generated using R programming by comparing the relative signal intensity of the selected genes in OSS and ESS.

3.3.3. Strain construction

All gene deletions were generated as previously described [51]. Briefly, cells carrying the plasmid pSIM5 were induced at 42℃ for 15 min, cooled immediately, and made electrocompetent. The competent cells were then transformed with PCR products carrying kanamycin gene and complementary regions upstream and downstream of the gene of interest. Recombinants were selected on LB agar containing kanamycin, and positive clones were verified by PCR and sequencing. Plasmid pZB-G was constructed by a sequence- and ligation-independent cloning (SLIC) method [56]. The *pgi*, *fbaA*, *gapA*, and *tpiA* genes were amplified with primers containing their native RBS sequences and regions homologous to their successive genes. Similarly, plasmid pZB was also amplified by PCR. The PCR products were then independently treated with T4 DNA polymerase for 20 min, and exonuclease activity was terminated by addition of dCTP. The vector and inserts were mixed (1:2 molar ratio) and transformed into electrocompetent cells. pZB-A was constructed by cloning *ascB* into the *Pst*I and *Avr*II sites of pZB. pPro-*ascB* was constructed by cloning *ascB* into the *Kpn*I and *Bam*HI sites of pPro29b.

3.3.4. Enzyme assay

Phospho-β-glucosidase assays were performed with OSS and ESS grown on minimal cellobiose medium. Cells were collected every 4 h and stored at –20°C. The cell pellet was dissolved in 100 μL of 100 mM sodium phosphate buffer (pH 7.0), 400 μL of 10 mM *p*-nitrophenyl-β-glucoside was added, and the samples were incubated at 37℃ for 40 min. The reaction was arrested by adding 1M sodium carbonate solution. The yellow color developed was quantified by measuring OD at 410 nm and 600 nm.

3.3.5. RT-PCR

Total RNA was isolated using Trizol reagent according to the manufacturer's procedure. cDNA was synthesized using random hexamers and MuLV reverse transcriptase. Expression of the target genes was quantified using gene-specific primers and Phusion® DNA polymerase. PCR samples (5 μL) were taken after 25, 27, 30, and 32 cycles and analyzed on 0.8% agarose gel. DNA concentration was determined based on the molecular weight marker.

3.4. Results and discussion

3.4.1. Global gene expression profiles of strains OSS and ESS

The specific growth rate of ESS on cellobiose was 2-fold higher than that of OSS (μ_{ESS} = 0.36 ± 0.05 h^{-1} and μ_{OSS} = 0.18 ± 0.04 h^{-1}). Thus, we hypothesized that comparison of transcriptomic polymorphisms would help in understanding the alteration in key regulatory pathways that could have contributed to the increased growth rate of ESS. Changes in mRNA levels of only a few genes were expected to have contributed to the fitness of *E. coli* on the non-native carbon source, cellobiose. Contrary to expectation, total of 804 genes exhibited a more than 2-fold difference in expression. These genes are distributed across a wide range of functional categories (Fig 3.1A & 3.1B), which were obtained by ADGO, and are listed in Table 3.2.

3.4.1.1. Decrease in stress response may contribute to the fitness

We hypothesized that constitutive expression of 2 cryptic operons to favor growth on a non-native carbon source (cellobiose) would trigger a stress response in OSS. As expected, the most abundant transcript in OSS corresponded to a stress response gene, *rmf* (ribosome modulation factor) (Fig 3.1C). The *rmf* gene is one of the three dominant genes (*rmf*, *ygaW*, *ghrA*) that are up-regulated in all types of stress response tested for *E. coli*, and it functions to decrease the overall translational rate by forming ribosome dimers [57]. It is not clear if the slow growth of OSS on cellobiose is due to elevated expression level of *rmf* or vice versa. The *rmf* mRNA level in ESS was 2.6-fold lower than in OSS, indicating a cause for increased fitness on cellobiose. The osmotic stress responsive gene *osmB* was reduced 10-fold in ESS. Similarly, the expression of *rpoS*, the master regulator of universal stress response, was decreased 2-fold in ESS. Consistent with this finding, several *rpoS*-regulated stress responsive genes such as *uspB* (tolerance to ethanol stress) and *ddpX* (dipeptidase) were also down-regulated. This evidence suggests a paradigm shift in OSS from stress due to non-native cryptic carbon source to active metabolism in ESS following adaptive evolution on cellobiose (Fig 3.1B).

3.4.1.2. Increased expression of flagellar biosynthesis

Flagellar organization and chemotaxis genes are highly expressed in cells that are actively growing on a poor-quality carbon source, a foraging strategy to search for a better carbon source [58]. Genes responsible for flagellar assembly are expressed 2- to 50-fold higher in ESS than in OSS (Fig 3.1B). Similarly, the expression of *trg*, the

chemotaxis receptor for glucose, galactose, and ribose, was 1.5-fold lower in ESS. Alterations in the chemotactic and locomotor behavior of ESS may also contribute to its increased fitness on cellobiose.

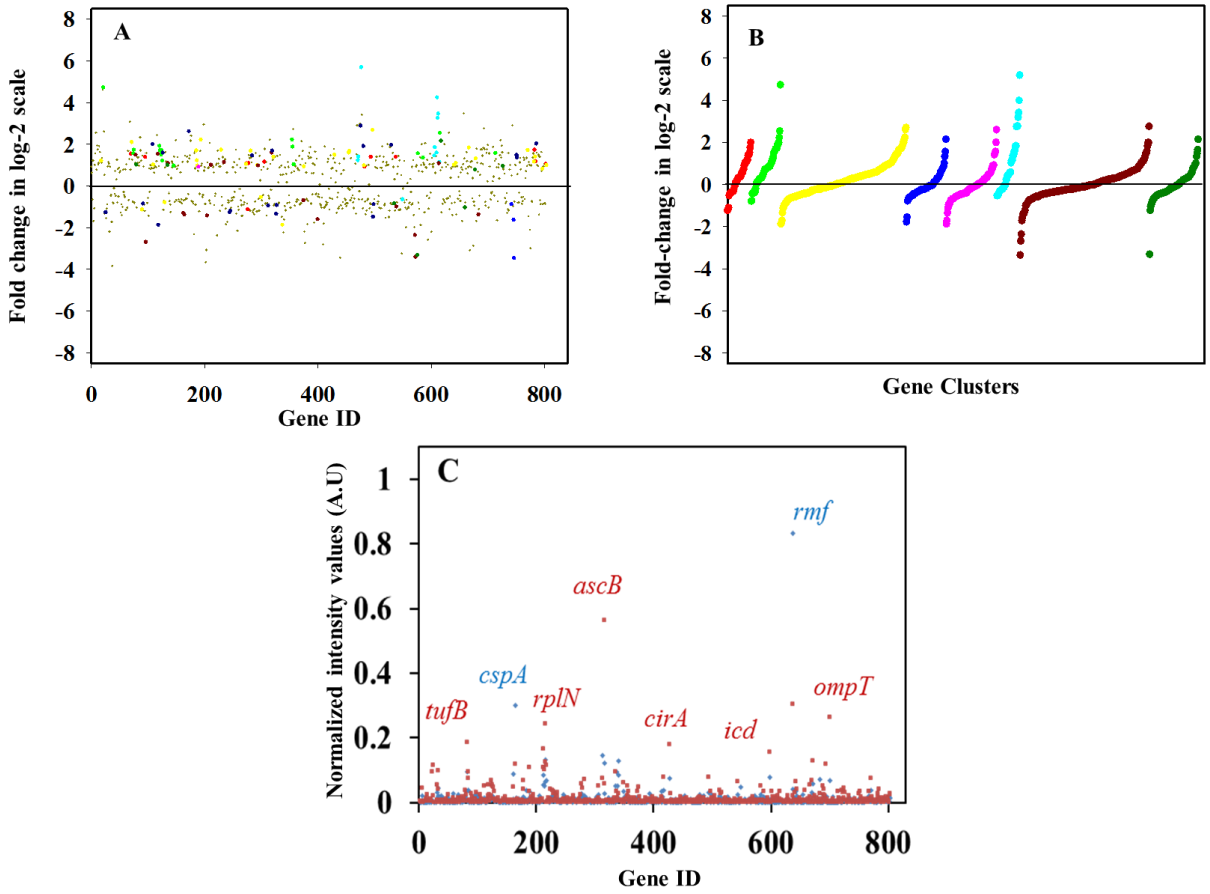


Fig. 3.1. (A) Scatter plot describing the distribution of fold-difference for the 800 differentially expressed genes in ESS and OSS. Fold-change is represented in log–2 scale. (B) Fold-change in the expression level of genes grouped using ADGO. Glucose metabolic pathway, red; nucleotide biosynthesis pathway, green; amino acid metabolism, yellow; lipid metabolism, blue; amine metabolism, pink; flagellar assembly and chemotaxis, cyan; transcription factors, brick brown; stress response, dark green. Fold-change is represented in log–2 scale. (C) Analysis of the most abundant mRNA among the 800 significantly altered genes in OSS and ESS. Signal intensity is normalized using the average signal intensity obtained from three independent trials and presented as arbitrary units (A.U). The most abundant mRNA in OSS (blue diamonds) corresponds to the universal stress response protein *rmf*. The most abundant mRNA in ESS (red squares) is that of *ascB*.

3.4.1.3. Amino acid and nucleotide metabolism

Diaminopimelate is a common metabolic intermediate in lysine and cell wall biosynthesis. Five of the 9 genes responsible for converting diaminopimelate to lysine were up-regulated in ESS. Genes related to lysine and methionine biosynthesis was also up-regulated in ESS indicating an active metabolism (Fig 3.1B). The F_1F_0 ATP synthase complex is encoded by *atpCHAGD* (F_1 complex) and *atpFEBI* (F_0 complex). These genes were induced 3-fold in ESS (except for *atpC* and *atpD* whose expression patterns were relatively unaltered). Many of the nucleotide biosynthesis genes were up-regulated in ESS. Transcription of *pyrI* and *pyrB*, encoding the subunits of aspartate transcarbamylase, which catalyzes the first step of the *de novo* biosynthesis of pyrimidine, was 25-fold higher in ESS. These results indicate active metabolism and fast growth. Actively growing cells generate more ATP, and hence, the up-regulation of genes necessary for ATP synthase. Similarly, nucleotides are needed to synthesize DNA and RNA and for catabolic processes in a growing cell. Up-regulation of genes related to *de novo* synthesis of nucleobases is another indication of fast growth in ESS.

3.4.1.4. Transcription factors

Unlike the metabolic genes, there may be a direct correlation between mRNA and protein expression of transcription factors, because the expression levels of these factors are lower than those of other proteins. In addition, expression of a particular transcription factor might in turn guarantee the expression of genes that are positively regulated by the same transcription factor. The mRNA levels of several transcription factors were 2- to 7-fold higher in ESS. For example, *flhD* and *fliA* are involved in regulating flagellar biosynthesis, and their expression levels were 4- and 7-fold higher in ESS, respectively. Accordingly, genes involved in flagellar biosynthesis were expressed at a higher level in ESS (see section 3.1.2. and Table 3.2). Expression of *xylR*, the transcriptional activator of the xylose operon, was also 1.5-fold higher in ESS. This result provides evidence for co-metabolism of xylose and cellobiose in ESS. In contrast, *crp* expression was reduced 3-fold in ESS, but none of the genes predicted to be regulated by *crp* were altered between ESS and OSS. This may be because these genes are controlled by several factors, including global regulators, gene-specific regulators, and metabolites.

3.4.2. Gene expression profile of central metabolic pathway of cellobiose

Bacterial porin, OmpT transports sugars and other small molecules from the exterior to the periplasmic space. *ompT* expression was high and constant in both ESS and OSS, indicating that periplasmic translocation of cellobiose may not be the rate-limiting step. One explanation for the increased growth of ESS on cellobiose could be the activation of central metabolic pathways through direct mutations in promoter sequences or transcription factors. Consistent with our hypothesis, 7 of the 41 major genes of CMP showed higher expression in ESS than in OSS. Furthermore, the specific growth rate of ESS on glucose was higher than that of OSS ($\mu_{ESS} = 0.47 \pm 0.01$ h^{-1} and $\mu_{OSS} = 0.39 \pm 0.04$ h^{-1}), indicating the possibility of strengthening CMP. Interestingly, no mutations were found within the promoter region of *pgi*, *ppc* and *aceA*, which shows 3-fold difference in expression between strain ESS and OSS. Genes of the upper glycolysis pathway, namely, *pgi*, *fbaA*, *gapA*, and *tpiA*, were induced (2.9-, 2.6-, 2.7-, and 2.6-fold, respectively) in ESS. However, individual or combinational overexpression of these 4 genes in OSS did not enhance cell growth on cellobiose. In fact, *fbaA* or *tpiA* overexpression caused severe growth arrest upon induction even in cells growing in glucose minimal medium, probably due to the toxicity caused by accumulation of phosphorylated glycolytic intermediates. The expression of multiple CMP genes at coordinated levels might increase overall metabolic pathway flux and the increased gene expression in EMP might be because of mutations in unknown transcriptional factors. Results of previous studies have suggested that elimination of stress would help restore active glycolysis and other central carbon metabolism [59]. Another explanation is that a decrease in the expression of stress-responsive genes (such as *rmf*) in ESS up-regulates the CMP.

Table 3.2. Representative functional categories of up/down-regulated genes obtained from ADGO

Functional category	2-fold higher expressed genes	2-fold lower expressed genes
Amine metabolism	*livF, ybaS, gcvT, glmU, gadB, gadA*	*astC, asnA, gabD, tdcA, astA, tdcE, gabT*
Amino acid metabolism	*gdhA, gadA, lysC, asd, metE, gadB, proX, dapD, hisG, gltD, aroG, livJ, serA, hisC, folE, betB, speD, proV, glnS, gcvT, speE, livK, valS, serC, dapF, asnS, gshB, thrA, gltX, livM, aspS, lysA, carB*	*gabT, tnaB, tdcE, astA, tdcA, metJ, argA, argC, gabD, tyrA, asnA, astC*
Flagellar assembly and chemotaxis	*fliC, flgE, flgB, flgD, fliA, flhD, flgL, flgG, fliF, flgC, fliG, flgN*	*trg*
Glucose metabolic pathway	*pdhR, aceE, pgi, maeA, tpiA, gapA, aceF, tktA, eno, pgk, talB, zwf*	*epd, treR*
Lipid metabolism	*lpxD, glmU*	*prpB, prpC, mhpC*
Nucleotide biosynthesis	*pyrI, pyrC, guaA, metE, pyrF, atpB, atpE, atpF, purD, atpH, guaB, purH, adk, atpA, purU, atpG, purM, gmk*	Nil
Stress response	*recA, xseB, ybfE, tufB, recF, betI, betB, proX, dinI*	*osmB, treR, dinG, rpoS, ddpX*
Transcription Factors	*rplD, gyrB, flgM, galR, hpf, betI, rpoB, rbsR, mprA, flhD, pdhR, fliA*	*pspA, cpxP, pspB, lrhA, yiaG, crp, cspE, cspA, treR, metJ, tdcA, yddM, yfeR, rpoS*

The pathway for cellobiose metabolism was predicted from mRNA expression levels (Fig 3.2A). Transcription of *ppc*, encoding the anaplerotic reaction connecting glycolysis with the TCA cycle, was higher in ESS. In general, expression of the glyoxylate pathway is absent in cells grown on glucose and active under glucose-limited growth conditions [60]. However, the glyoxylate pathway gene *aceA* is the most abundant transcript after *rmf* in OSS. After adaptation, *aceA* expression was reduced by 1.7-fold in ESS (Fig 3.2A &3. 2B). Both OSS and ESS strains expressed no transcripts corresponding to *sucAB*, indicating that the TCA cycle is not fully active in these strains.

3.4.3. Genetic features of *ascB* and *chbF* and cell growth

ascB and *chbF* encode a phospho-β-glucosidase of size 54kDa and 50kDa, respectively that hydrolyzes intracellular cellobiose-6-phosphate and are the second and fifth genes in the *asc* and *chb* operons, respectively. In engineered ESS and OSS, both these operons are expressed from a synthetic constitutive promoter CP12 [26]. Surprisingly, the most abundant mRNA expressed in ESS corresponds to the *ascB* gene, compared to the stress responsive gene *rmf* in OSS. We hypothesized that the change in *ascB* expression might have a significant impact

on cellobiose metabolism, as its expression was induced 5-fold in ESS (Fig 3.2A & 3.2B). The increased *ascB* expression level was confirmed by RT-PCR. Therefore, it is possible that the *ascB* gene has its own promoter that may be positively controlled by cellobiose metabolism. In our previous study, no significant difference was observed in the β-glulcosidase enzyme activities between ESS and OSS grown in LB broth [26]. Enzyme activity was measured in ESS and OSS grown on cellobiose. As expected, the enzyme activity was several-fold lower in OSS than in ESS (Fig 3.3C). However, no significant difference was observed in enzyme activity between ESS and OSS grown in glucose minimal medium (data not shown). Thus, increased expression of *ascB* in ESS may be inducible in the presence of cellobiose, a factor that strengthens our assumption of a putative promoter for *ascB*.

Cellobiose metabolism was rescued in OSS (equivalent to that of ESS) when *ascB* was overexpressed from P_{zt1} on a low-copy plasmid (Fig 3.3A). A similar effect was observed when *ascB* was expressed from a stronger promoter P_{pro} on a medium-copy plasmid (data not shown). Even though cellobiose metabolism was not completely rescued when *chbF* was overexpressed in OSS, there was a marked increase in cell growth. These results indicate that β-glulcosidase activity is the rate-limiting step, and its overexpression facilitated rescue of cellobiose metabolism (Fig 3.3A). Similarly, when *ascB* was deleted from ESS, a rapid loss of growth was observed on cellobiose, whereas *chbF* deletion did not affect cellobiose metabolism in ESS (Fig 3.3B). In addition, no growth on cellobiose was observed when *chbBAC* was deleted in ESS. This result indicates that β-glucosidase, AscB, and the cellobiose transporter ChbBCA enhance cellobiose metabolism.

Previously, we have reported adaptation of strains expressing only the *chb* operon (OSS-C) or both *chb* and *asc* operons (OSS) from constitutive promoters. Even after 75 days of adaptation, OSS-C could not attain efficient cellobiose metabolism despite the fact that the *chb* operon is the major contributor to the process. Efficient cellobiose metabolism was obtained only in OSS in which the *asc* operon was also expressed. Strain OSS-A expressing only the *asc* operon from a synthetic constitutive promoter did not grow on cellobiose-minimal medium. Overexpression of neither *ascB* nor *chbBAC* conferred the cellobiose-metabolizing phenotype in OSS-A. These results confirm that the combined and coordinated expression of all genes of the *chb* and *asc* operons is necessary for efficient cellobiose metabolism.

Figure 2

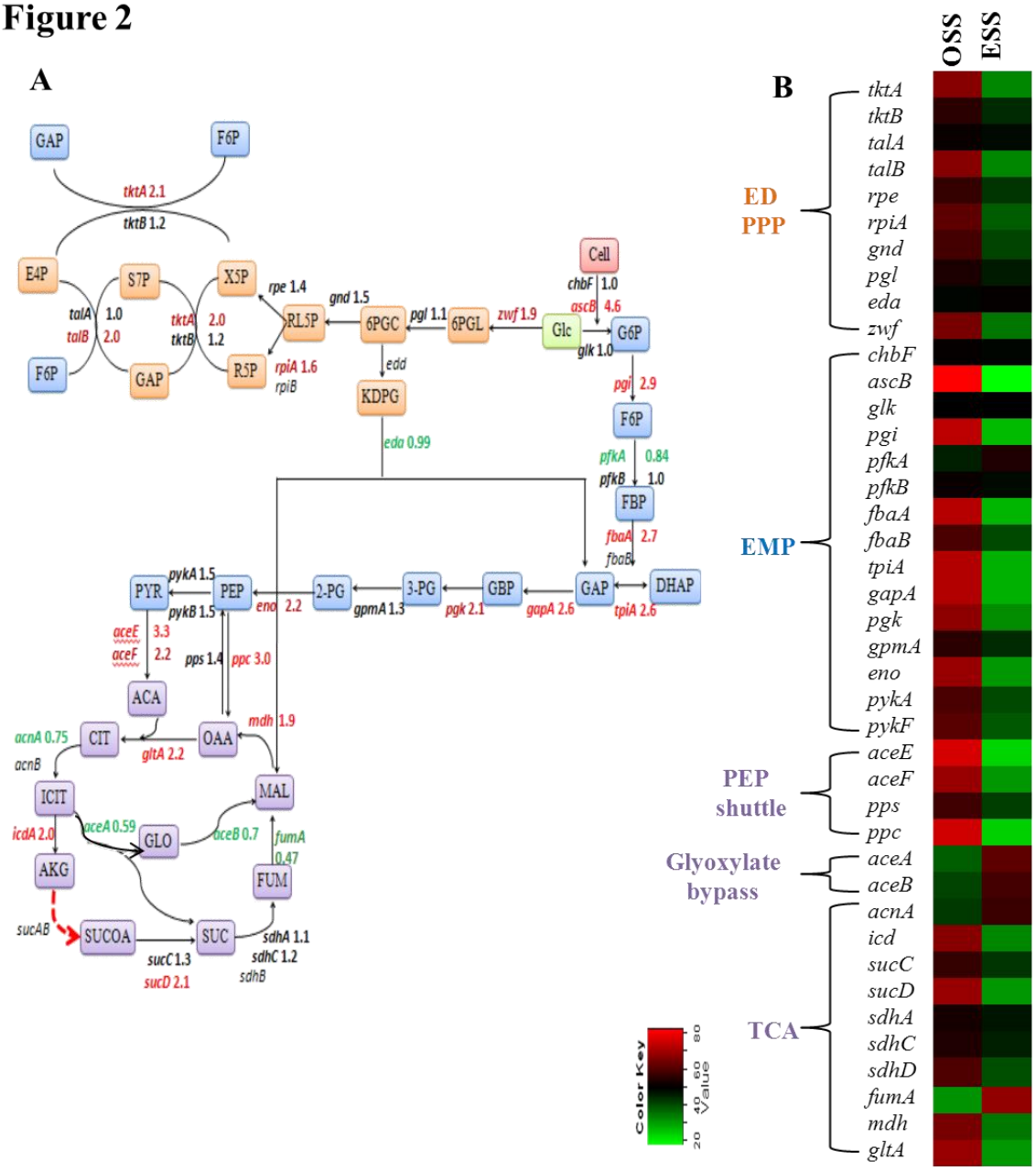


Fig. 3.2. Comparative transcriptome profiling of OSS and ESS. (A) Numbers are the ratio of signal intensity of each ESS gene to its expression in OSS. Genes in red are significantly upregulated and those in green are significantly downregulated in ESS. Dotted arrows represent the pathway corresponding to missing transcript in the microarray. Genes in black were unchanged or not detected with accuracy in the microarray. (B) Heatmap to compare the relative mRNA abundance of the central metabolic pathway in OSS and ESS. The first column

corresponds to OSS and the second column to ESS. ED, Entner-Doudoroff pathway; PPP, Pentose Phosphate Pathway; EMP, Embden-Meyerhof-Parnas pathway; PEP, Phosphoenolpyruvate; TCA, tricarboxylic acid cycle.

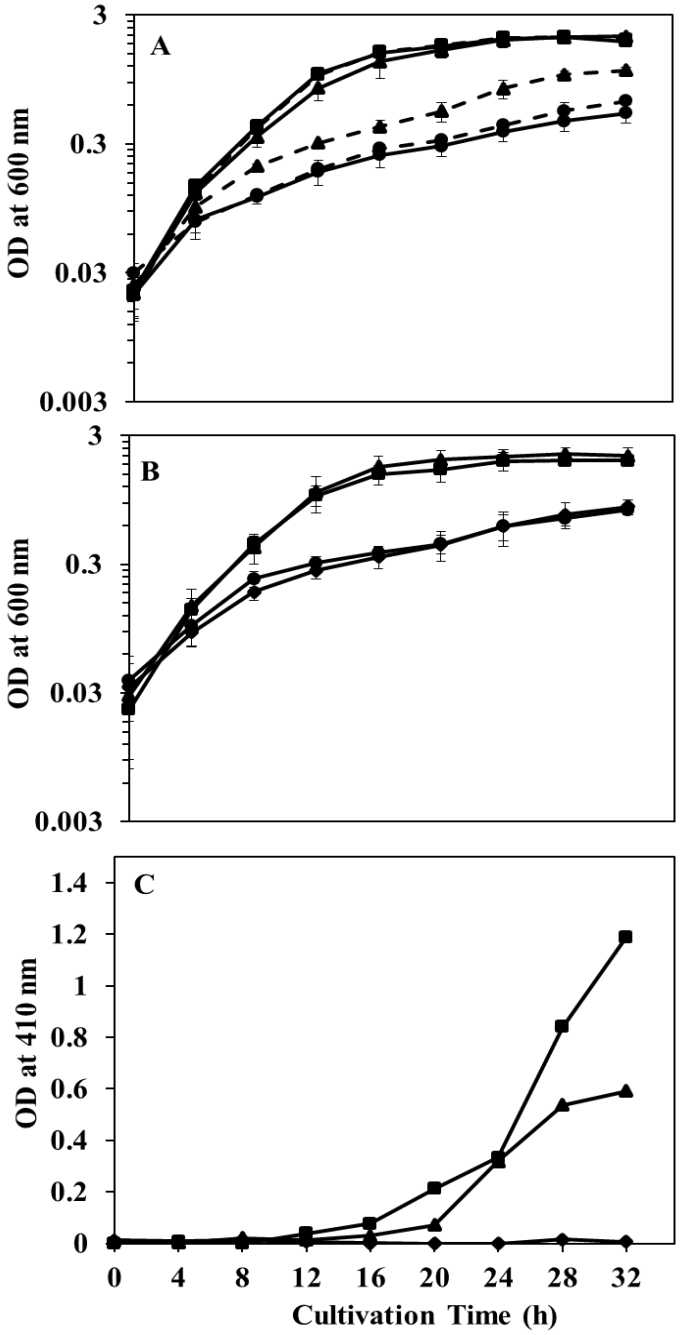


Fig. 3.3. Analysis of efficiency of *ascB* to favor efficient cellobiose metabolism. (A) *ascB* was overexpressed from P_{zt1} in a low-copy plasmid. Dashed lines represent strains grown without the inducer (tetracycline). Solid lines correspond to strains grown in the presence of 2 ng/mL tetracycline. OSS, closed circle; ESS, closed square;

OSS-pZB-A, closed triangle. Error bars represent the standard deviation of triplicate experiments (B) Deletion of *ascB* in OSS to reanalyze the effect of *ascB* on cellobiose metabolism. OSS, closed diamond; ESS, closed square; ESS-F, closed triangle. ESS-B, closed circle. Error bars represent the standard deviation of triplicate experiments. (C) β-glucosidase production and stability was determined by using PNPG as a substrate. Approximately 10^9 cells were collected at different time points and PNPG activity was measured as indicated in the Materials and Methods. OSS, closed diamond; ESS, closed square; OSS-pZB-A, closed triangle. Data are representative of a triplicate experiment.

3.5. Conclusions

Transcriptome profiling was used to determine the rate-limiting steps in cellobiose metabolism. Our results provide evidence that the *asc* operon and its enzyme may be as important as the *chb* operon for efficient cellobiose metabolism. Promoter engineering and transcriptome profiling revealed the importance and superiority of the *asc* operon for cellobiose metabolism. This study could also be used to decipher the crypticity associated with these operons. Even though transcriptome analysis indicates significant alterations in gene-expression patterns, it cannot be used to characterize the molecular underpinnings leading to expression polymorphisms. It is still not clear if the increase in *ascB* expression enabled efficient cellobiose metabolism and a reduced stress response or vice versa. Even after we establish that increasing *ascB* expression level increases fitness on cellobiose, further genomic, proteomic, and metabolomic studies are needed to validate the reasons for differences in *ascB* expression between OSS and ESS.

Chapter 4

Genotypic and phenotypic characterization of strain ESS

(Novel functions and regulations of cryptic cellobiose operons in *Escherichia coli*)

4.1. Abstract

In this study, we have characterized the two mutations observed in the efficient cellobiose metabolizing strain: duplication of RBS of *ascB* gene, (β-glucosidase of *asc* operon) and nonsense mutation in *yebK*, (an uncharacterized transcription factor). Mutations in *yebK* plays a dominant role by modulating the length of lag phase relative to the growth rate of the strain when transferred from a rich medium to minimal medium. Mutations in *ascB*, on the other hand, are specific for cellobiose and helps in enhancing the specific growth rate on cellobiose. Here, we show that engineering transporter and β- glucosidase alone may not be sufficient and changes in the endogenous regulatory circuit is mandatory for achieving efficient cellobiose metabolism. The transcription factor *yebK* thus, helps in establishing a functional coupling of cellobiose metabolism to *E. coli* physiology.

4.2. Introduction

Previous studies for understanding the cryptic cellobiose metabolism was based on characterizing stress-directed mutations in the *chb* and *asc* operons [61]. However, these studies could not establish host factors (other than the genes of the *chb*/*asc* operons) that control the cryptic cellobiose metabolism [61]. Here, we describe the functional benefits of the activated *chb* and *asc* operons and decipher the regulatory changes that occurred during adaptation on cellobiose minimal medium. In bacteria, the transcriptional regulatory network plays a significant role in helping bacteria to adapt to the nutrients in the medium. Detailed knowledge of the endogenous regulatory network that shows differential response to the activated cryptic genes is thus important to rewire metabolic pathways for efficient cellobiose metabolism. While previous studies characterized the stress-induced mutations that allowed growth on cellobiose [61], here we characterize the adaptation induced mutations that helps in enhancing cellobiose metabolic ability. This study will thus help in understanding the regulatory changes and factors that are needed for the functional coupling of the host physiology to the activated cryptic cellobiose metabolism.

4.3. Materials and methods

4.3.1. Bacterial strains and media composition

All strains and plasmids used in this study are listed in Table 4.1. Bacteria were cultured at 37℃ in Luria Bertani broth (LB) or minimal medium supplemented with cellobiose. Strains carrying temperature-sensitive plasmids were grown at 30℃. Media were supplemented with suitable antibiotics (30 µg chloramphenicol/mL and 100 µg ampicillin /mL). For long-term storage, cells were maintained as 20% glycerol stocks at –80℃.

M9 minimal medium, supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 4 g sugar/L, was used to characterize the cell growth rate of the modified strains. In these tests, overnight cultures grown in LB were collected, washed once with M9-salts and suspended to a final OD of 0.05 in 50 mL of M9 medium supplemented with the test sugar in a 250-mL flask, and the cultures were incubated at 37℃ with rotation at 200 rpm. Cell growth was monitored by measuring the optical density at 600 nm (OD_{600}) every 3 hours with a Biochrom Libra S22 spectrophotometer. The lag phase was calculated as the time taken for the cells to reach an OD of 0.2. For gentiobiose as a carbon source, 1% of overnight grown culture (in LB) was inoculated into M9-gentiobiose minimal medium and were grown on micro titer plates (200µL) and OD_{600} was measured every 10 minutes using TECAN plate reader at 37℃ with rotation at 88 rpm.

Table 4.1. Bacterial strains and plasmids used in this study

Strains/plasmids	Description/genotype	Reference/source
Strains		
E. coli MG1655	Wild type	[37]
MG1655/Δ*yebK*::*frt*	MG1655 with *yebK* gene deleted	This study
Enterobacter cloacae	ATCC 13047	KCTC
OSS	MG1655 with CP12chb plus CP12asc	[18]
ESS	OSS adapted in cellobiose for 30 days	[18]
OSS-*yebK**	OSS with nonsense mutation in *yebK*	This study
OSS-*ascB**	OSS with 10bp inserted in the RBS region of *ascB*	This study
OSS-*yebK*/ascB**	OSS with both mutations in ESS	This study
OSS-Δ*eda*::*frt*	OSS with *eda* gene deleted	This study
OSS-Δ*edd*::*frt*	OSS with *edd* gene deleted	This study
OSS-*yebK**/Δ*edd*::*frt*	OSS with nonsense mutation in *yebK* and *edd*gene deleted	This study
OSS- P2E5	OSS with RBS of *ascB* mutated through MAGE	This study
Plasmids		
pSIM5	λ-Red recombinase expression plasmid and temperature-sensitive replication	[62]
pCP20	Yeast FLP recombinase gene controlled by cI repressor and temperaturesensitive replication.	[39]
pKD13	Template plasmid for gene disruption. The kanamycin resistance gene is flanked by FRT sites. oriR6K-gamma origin requiring the pir+ *E. coli*.	[63]
pKD13-SacB	Modified form of pKD13 with kanamycing gene co-expressed with *sacB* gene	This study
pLcell-1	pL31b+ vector expressing the ECL-00047 to ECL-00051 gene cluster from *Enterobacter cloacae*	This study
pET31b+*yebK*-6his	pET31b+ plasmid with *yebK*-6his	This study
pET31b+*yebK**-6his	pET31b+ plasmid with *yebK**-6his	This study

4.3.2. Whole genome re-sequencing

For whole genome re-sequencing, DNA was isolated from both OSS and ESS using GeneAll DNA isolation kit. Library construction and DNA sequencing was performed by Macrogen Company (Korea) on an illumina Hiseq2000 platform. *E. coli* K12 MG1655 (NC_00913.2) sequence was used as a reference sequence.

4.3.3. Strain construction

Gene deletion was performed using the λ-Red recombination system, as described previously [63]. Briefly, the kanamycin cassette of pKD13 was amplified using the deletion primers containing 50bp homology to the target genes. Cells carrying the λ -Red system, under the control of the P_L promoter (pSIM5), were induced at 42℃ for 15 minutes, made electro competent, and then transformed with the PCR product. Transformant colonies carrying the desired modification were directly selected on LB agar plates supplemented with kanamycin. The kanamycin cassette was then cured using the FLP recombinase expressed from pCP20 plasmid. Genomic DNA was isolated from the transformants, and the target region was PCR amplified and sequenced to confirm site-specific insertion. The primers used for strain construction are listed in Table 4.2.

OSS-*yebK** was constructed using a scar-less deletion method. The kanamycin cassette in the pKD13 plasmid was modified to include *sacB* gene. The *kan-sacB* cassette was amplified with primers carrying 50bp homology to the target gene and gene deletion was performed using the λ-Red recombination system as described above. Sucrose sensitive strains were then selected by negative screening on LB sucrose plates (containing no salts). For *yebK**, the *yebK* gene was amplified from ESS and transformed to the sucrose sensitive OSS-Δ*yebK*::*kan-sacB* strain essentially as described above. The transformants were selected on LB-sucrose plates and confirmed by DNA sequencing.

OSS-*ascB** strains were constructed by single stranded oligo-mediated recombineering. Oligos were designed for the insertion of 10 nucleotides that was found duplicated in the ESS strains. Genome engineering was performed for 4 cycles using 2μM oligo per cycle. 96 colonies were randomly picked and screened on minimal cellobiose medium to identify clones with 10bp inserted in the *ascB* region. Clones were screened without an enrichment to avoid off target mutations. The efficiency of insertion of the 10bp was 5%.

4.3.4. Plasmid construction

The plasmid pLcell-1 was constructed by amplifying the ECL-00047 to ECL-00051 gene cluster from *Enterobacter cloacae* using primers, cell1_FP and cell1_RP. The resulting PCR product was digested with *XhoI*, and *XbaI* site and cloned into the plasmid pL31b+.

Table 4.2. Primers used in the study. Sequence in red indicates the duplicated nuclotide observed in strain ESS. The underlined sequence indicate the restriction site used for cloning. The randomized RBS sequence is indicated as NNNNNN.

Primer Name	Sequence
*chbB*_RBS	AACAAAACAGATAAATGTGTTTCTTTTCCATAAAACTGCCCTNNNNNNCGATTATCTGTCAGCCAGACACTCCGCAAGCCTTAACCTGCT
*chbC*_RBS	CCTTTTCAAGCGATGCAATAACATTACTCATAGAAAAATACCNNNNNNAACCGCAATTTAAATATTGCGGTATTGATTTATGAAATAACT
*chbA*_RBS	CGTTTGCGTATCGGGAATGTTATCGAGATCCATCATACATCGNNNNNNTCTTTTCTTACCGGCACGATTACCCGTACCGGCATCGATTAA
*chbF*_RBS	CACCAATAGTGACGACTTTTAATTTCTGGCTCATAATTTCTCNNNNNNGTACAGAATACTGATATCTGGCATATCTGCCCCCCCGGACAT
*ascF*_RBS	ACTAGTCGGCCAAAATGATATAATACCTGAGTACTGTTCACANNNNNNACAGCTATGGCCAAAAATTATGCGGCGCTGGCACGCTCGGTG
*ascB*_RBS	TCAGAGCGTACAACCGACCGTCGCCAAAGAAGTAAGTCTTAANNNNNNATGAAAATGTCAGTATTTCCAGAAAGTTTTTTATGGGGCGGC
*ascB*_ESS_RBS	TACAACCGACCGTCGCCAAAGAAGTAAGTCTTAATT**GAGGATGAAA**GAGGATGAAAATGTCAGTATTTCCAGAAAGTTTTTTATGGGGCG
CP_Direct	CATATACAAGTTTATTCTTGACACTAGTCGGCCAANNNNNNNNNNTACCTGAGTACTGTTCACACAGGAAACAGCTATG
CP_Complement	CATAGCTGTTTCCTGTGTGAACAGTACTCAGGTANNNNNNNNNNTTGGCCGACTAGTGTCAAGAATAAACTTGTATATG
*yebK*_FP	TTTCTTTCAGTGCGGAAATCGTCATTACCCGTGAGTCTCTTTACATCATGGTGTAGGCTGGAGCTGCTTCG
*yebK*_RP	GTATAAGATTAGGACAGTGACAGTCGTTTTTAGCGATCGTCACTTAAATTATTCCGGGGATCCGTCGACC`
*eda*_FP	GCCCGATCCTCGATCGGGCATTTTGACTTTTACAGCTTAGCGCCTTCTACGTGTAGGCTGGAGCTGCTTCG
*eda*_RP	TCACTTTTTAAGACGACAAATTTGTAATCAGGCGAGAGAAAACTCTGATGATTCCGGGGATCCGTCGACC
edd FP	TTCTCTCGCCTGATTACAAATTTGTCGTCTTAAAAAGTGATACAGGTTGCGTGTAGGCTGGAGCTGCTTCG
edd RP	CGCGTTGTGAATCATCCTGCTCTGACAACTCAATTTCAGGAGCCTTTATGATTCCGGGGATCCGTCGACC
*cell*1_FP	TCATCTAGATTAACGCTTACGCACATCCG
*cell*1_RP	AGCTCGAGTTATCTCTGATATAAATGAA
ascB-RT-FP	CGCCAACCAGTCTGAAGGTGCGTT
ascB-RT-RP	CCGGAGAACGGGCTATGCAACATA
ascF-RT-FP	GTTGATAACATCTCGGCGGTCACG
ascF-RT-RP	CCAGATACCAATCGGGCCAATCA
ascFB-RT-FP	TCTTCGATCCGGCGAATCCAATGA
ascFB-RT-RP	GGAAATACTGACATTTTCAT
ascF-5'-GSP2	TGACGACTTCCTGAAAGGCTTGTGA
ascB-5'-GSP2	AAAGCAGGTTCTGGCGTAGCGGCT
Q_T	CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGCTTTTTTTTTTTTTTTTTTTTT
Q_0	CCAGTGAGCAGAGTGACG
Q_I	GAGGACTCGAGCTCAAGC
YebK_EMSA_FP	GGGTAAAAATCTGACACTGATCATG
YebK_EMSA_RP	TCCGCACTGAAAGAAATCGAAATGC

4.3.5. Phenotype microarray (PM)

The PM assay was performed using Biolog Inc essentially as per the instructions [64]. Briefly, ESS strains were plated out on LB agar prior to starting the assay. Each colony was suspended in inoculating fluid at the desired cell density and added to each well of the 10 Biolog plates containing different test compounds. The plates were then placed in the OmniLog reader (Biolog), and incubated for 72 h monitoring the respiration every 15 minutes.

The OmniLog reader converts the pixel density in each well to a signal value reflecting cell growth and dye conversion. Wild type MG1655 was used as a control in these experiments.

4.3.6. Multiplex Genome Engineering

Oligonucleotides were designed to randomize the RBS regions of *chbB*, *chbA*, *chbC*, *chbF*, *ascF* and *ascB* genes and the promoter regions of *chb* and *asc* operon. OSS strains were transformed with the plasmid pSIM5. Multiplex genome engineering was performed as described elsewhere but with a simple modification [65]. 7 oligos were mixed at a concentration of 0.3µM per oligo. Following every two cycles of genome engineering, strains were enriched on cellobiose to an OD600 of 0.5 and the enriched population was subjected to further MAGE cycle. The MAGE cycles were repeated 8 times with enrichment on cellobiose between 2 MAGE cycles. Approximately 282 clones were scored for efficient growth on cellobiose in 2-mL volume 96 well plates. The top scoring strains were analyzed again on test tubes for efficient growth on cellobiose. The entire *chb* and *asc* operons of the efficient cellobiose metabolizing strain were sequenced to identify the RBS regions mutated through genome engineering. The RBS strength was predicted using the RBSDesigner program [66].

4.3.7. Gene Specific RT-PCR and 5' RACE

Total RNA was isolated (using Qiagen RNA kit) from mid-log phase culture of OSS and ESS grown on cellobiose minimal medium as described above. Gene specific RT-PCR was performed using the Bioneer one step RT-PCR premix as described by the manufacturer using 0.1µg of total RNA. Samples were withdrawn at 20, 22, 24, 26, 28, and 30 PCR cycles and analyzed on 1 % agarose gel. 5' RACE PCR was performed as described previously [67]. The resulting PCR product was sequenced at Macrogen. Inc using the *ascF*-5'-GSP2 and *ascB*-5'-GSP2 primers respectively for *ascF* and *ascB* RACE.

4.3.8. YebK purification and EMSA

YebK or YebK* was expressed with 6His tags from pET31b+ plasmid. Proteins were purified as described before but with a simple modification [68]. In the last step, the purified proteins were maintained on Tris buffer as HEPES buffer led to the immediate precipitation of the protein. The template for EMSA was PCR amplified from MG1655 genomic DNA using the primer pairs: YebK_EMSA_FP and YebK_EMSA_RP. 150 ng of the DNA was incubated with proteins in the concentration ranging from 0 to 1 µM in binding buffer (16 mM Tris pH7.5, 3 mM $MgCl_2$, 30 mM NaCl, 0.0065 % Triton-X and 0.033 mg/mL BSA) for 1 hour at 37° C. When indicated, KDPG was substituted at a final concentration of 3 mM. The reaction was analyzed on 5% nondenaturing polyacrylamide gel for YebK or 7% non-denaturing polyacrylamide gel for YebK*.

4.4. Results and discussion

4.4.1. ESS utilizes a variety of substrates

Phenotype Microarray (PM) is used to monitor the phenotypic changes in the mutant bacteria (compared to wild type) across a variety of conditions simultaneously. The strain ESS (Evolved Synthetic Strain) is one of the efficient cellobiose metabolizing strains of *E. coli* constructed by replacing the cryptic promoters of *chb* and *asc* operons with constitutive promoters and adaptive evolution on cellobiose. Screening strain ESS on PM plates could help envision the impact of the activated cryptic operons and the mutations gained during adaptation on different nutrient conditions.

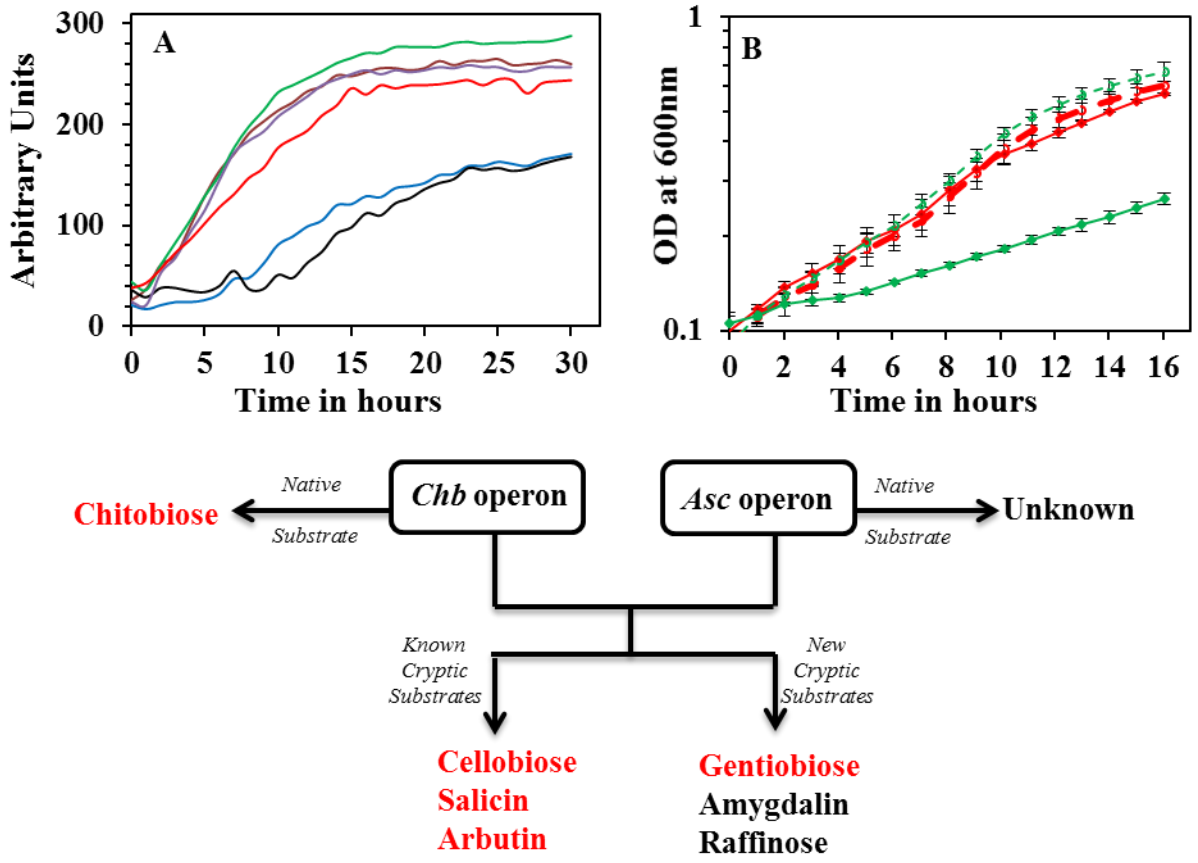


Fig. 4.1. (A) Measure of NADH concentration as a function of respiration and growth on different substrates using Biolog PM plates. *Purple*- salicin; *red*- arbutin; *dark red*- cellobiose; *green*- gentiobiose; *blue*- amygdalin; *black* – raffinose. (B) Growth of ESS (*solid line*) and OSS (*dashed line*) on cellobiose (*dark red*) or gentiobiose (*green*) minimal medium. The OD_{600} was measured every 10 minutes using TECAN microplate reader. Error bars indicate

the standard deviation of experiments performed in triplicates. (C) Scheme for the broad substrate range for metabolism offered by the combined activity of *chb* and *asc* operons as identified in this study. Substrates efficiently metabolized by strain ESS are indicated in bold faces.

The activated cryptic operons favored the strain ESS to grow on different non-native carbon sources (other than cellobiose) including salicin, arbutin, amygdalin, raffinose and gentiobiose (Fig 4.1A and 4.1B) indicating that the combined activity of the *chb* and the *asc* operons could enhance the metabolic ability of *E. coli* even beyond cellobiose. Interestingly, both OSS and ESS strains had higher (and comparable) growth rate on gentiobiose whereas only strain ESS exhibited higher cellobiose metabolism (Fig 4.1B and 4.1C). This might suggest that the *chb* and *asc* operon may have higher metabolic potential towards gentiobiose than cellobiose. Hence, it is necessary to characterize all mutations in strain, ESS in order to decipher the regulations of the cryptic *β*- glucoside metabolism.

It is also interesting to know that the genes encoded by the *chb* and the *asc* operons have higher metabolic ability towards gentiobiose. Even though *chbF* was previously reported to be active on phosphorylated gentiobiose, K_m for hydrolysis of gentiobiose was very high (12.5 mM for gentiobiose and 1.3 mM for cellobiose) [69]. Efficient gentiobiose metabolism by strain OSS provides a new challenge to the crypticity of *chb*/*asc* operons. The crypticity of *chb* and *asc* operons were attributed to the tendency to exclude the metabolism of toxic *β*- glucosides present in nature [10]. Here we show that strain ESS (with activated *chb* and *asc* operon) could also metabolize the toxic *β*- glucoside, Amygdalin albeit slowly. A combination of recombinant DNA technology (through promoter engineering), adaptive evolution and high-throughput screening helps in addressing the function of cryptic genes of *E. coli* in greater depths than with the conventional genetics.

4.4.2. Mutations pertained to cellobiose metabolism

Comparative whole genome re-sequencing reveals two mutations in strain ESS. The first mutation leads to the duplication of 10*nt* (GAGGATGAAA) above the *ascB* gene of the *asc* operon. The second mutation was a nonsense mutation on a previously uncharacterized transcription factor, *yebK* resulting in the expression of a truncated protein (79 amino acids) containing only the DNA binding domain. We then characterized the role of independent mutations on cellobiose metabolism. Allelic replacement of each of the two mutations independently or in

combination in strain OSS leads to enhancing the growth rate on cellobiose in the following order: OSS-*yebK***ascB**>*ascB**>*yebK**>OSS indicating that both of these mutations are beneficial for growth on cellobiose (Fig 4.2A and 4.2B). The final strain, OSS-*yebK***ascB** had a specific growth rate similar to the strain ESS when grown on cellobiose minimal medium indicating that all possible mutations related to cellobiose metabolism was deciphered.

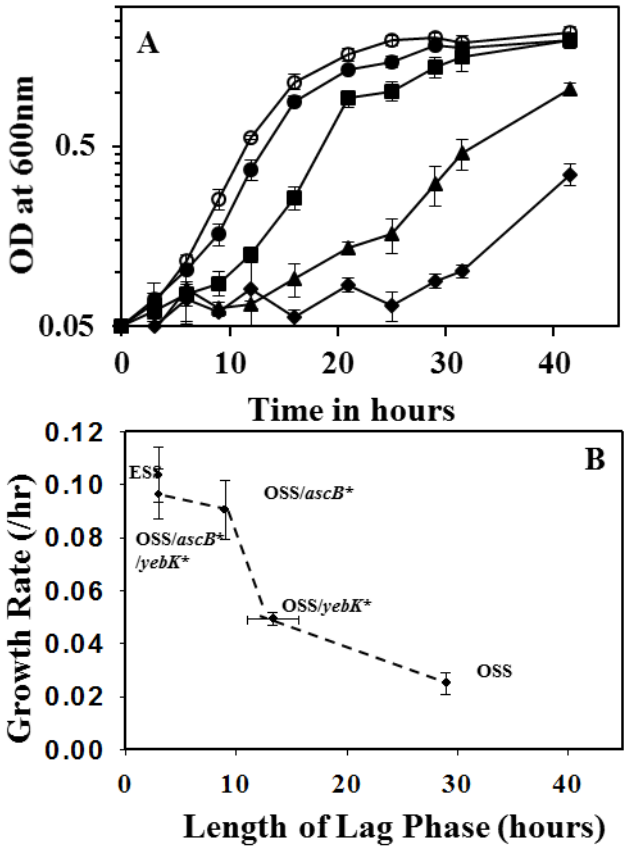


Fig. 4.2. (A) Comparison of growth of *E. coli* strains on cellobiose minimal medium. OSS; *closed diamond*, OSS-*yebK**; *closed triangle*, OSS-*ascB**; *closed square*, OSS-*yebK**/*ascB**; *closed circle*, ESS; *open circle*. (B) Correlation between the length of lag phase upon transfer from rich medium to cellobiose minimal medium and the specific growth rate on cellobiose. Error bars indicate the standard deviation of experiments performed in triplicates.

4.4.3. Role of AscB* in cellobiose metabolism

The 10nt above the start codon of *ascB* gene was duplicated in the strain, ESS. The duplicated nucleotides would

likely be related to enhancing the translational or transcriptional efficiency of *ascB* gene. Prediction of the RBS strength of *ascB* of strains, ESS and OSS revealed that the duplicated nucleotides doesn't affect the RBS strength or the translational efficiency but changes the length of the spacer sequence (Table 4.3). Hence, the duplicated nucleotides might play a different role (including transcriptional regulation or mRNA secondary structure change) other than to control the translational efficiency of *ascB*.

Table 4.3. Predicted RBS site (shown in bold face) of different genes related to cellobiose metabolism. The start codon is represented in red. The spacer length and efficiency are predicted using the RBSDesigner program. The second RBS of strain ESS is underlined.

Gene	Predicted RBS sequence	Spacer length	Spacer efficiency
chbB	GATAATCG**TCGATG**AGGGCAGTTTT**ATG**	11	0.568
chbC	TTAAATTGCGGTT**TTTAAG**GGTATTTTTCT**ATG**	11	0.568
chbA	CGTGCCGGTAAGAAAAGA**GAGGAA**CGATGT**ATG**	6	0.8591
chbF	TATCAGTATTCTGTAC**TGAAGG**GAGAAATT**ATG**	8	1
ascF	GTACTGTTCACA**CAGGAA**ACAGCT**ATG**	6	0.8591
ascB	AGTCTTAATT**GAGGAT**GAAA**ATG**	4	0.233
ascB of ESS	AGTCTTAATT**GAGGAT**GAAA**GAGGAT**GAAA**ATG**	14	0.219
ascB of P2E5	AGTCTTAA**AAGGAG**ATGAAA**ATG**	6	0.859

We previously reported through transcriptomic analysis that the *ascB* expression level was 5-fold higher in ESS compared to OSS [70] without any changes in *ascF* expression level (Fig 4.4A) despite the fact that *ascF* is the first gene of the operon and both *ascF* and *ascB* are organized as a single operon (Fig 4.3C). It could be possible that there is an additional promoter within the *asc* operon regulating the *ascB* gene. Hence, 5' RACE PCR was used to map the 5' region from *ascF* or *ascB* mRNA. The length of 5' RACE PCR product was 262bp as expected and the TSS of *ascF* mRNA was mapped to the TSS introduced along with the constitutive promoter (Fig 4.3E). However, the size of 5' RACE PCR from *ascB* was 553bp which is shorter than the expected 1995bp for full length *ascFB* transcript (Fig 4.3E). Upon sequencing the RACE PCR product, the TSS was mapped to 1419bp from *ascF* start

codon and was about 38bp above the mutation observed in strain, ESS. These results indicate that the modifications observed in the intergenic region of *ascB* in ESS would play a potential role in enhancing the 5' stability of the *ascB* transcript rather to enhance the RBS strength.

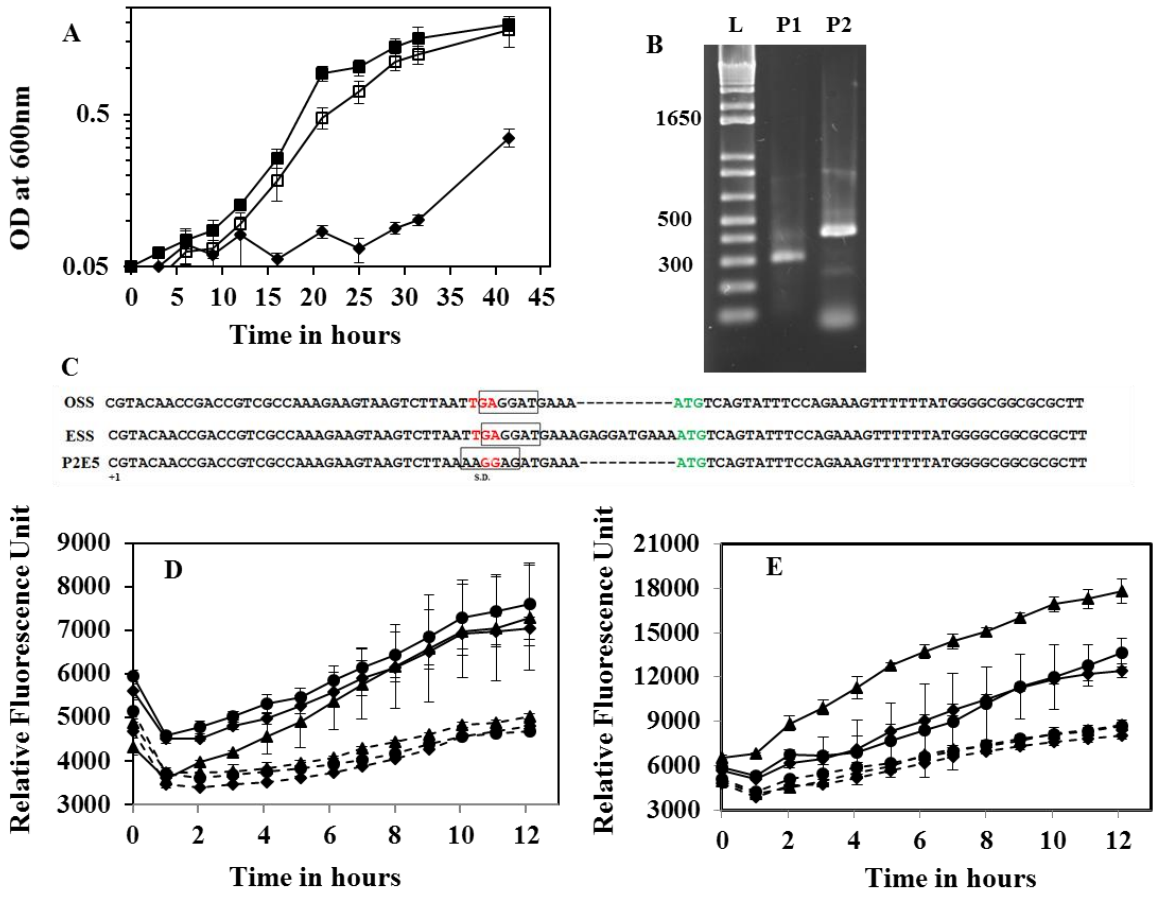


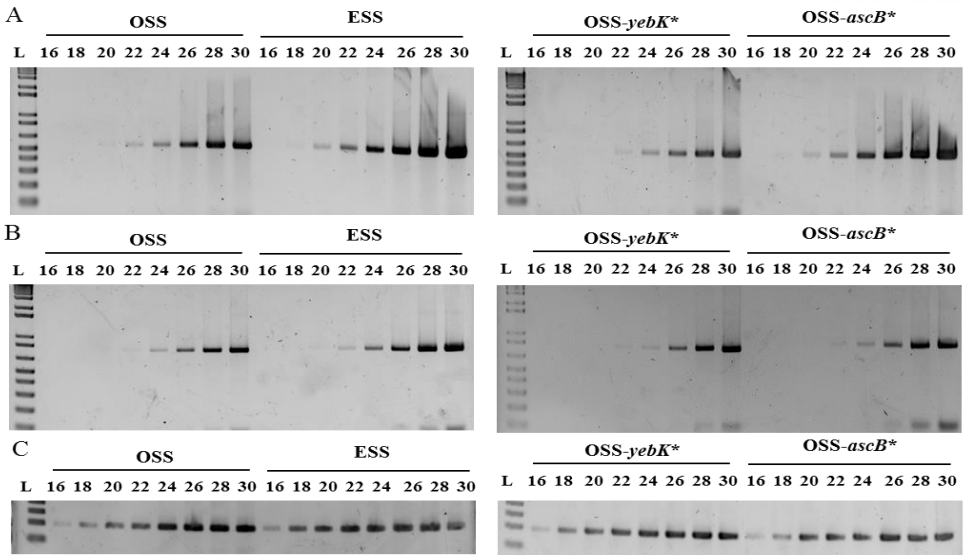
Fig. 4.3. (A) Comparison of growth on cellobiose. OSS; *closed diamond*, OSS-*ascB**; *closed square*, OSS-P2E5; *open square*. Error bars indicate the standard deviation of experiments performed in triplicates. (B) 5' RACE PCR to map the location of TSS for *ascF* and *ascB*. P1 corresponds to the transcript amplifiable by *ascF*_RACE_RP and is expected to be around 262bp; P2 corresponds to the transcript amplifiable by *ascB*_RACE_RP and is expected to be around 1995bp. The RACE PCR product of *ascF* corresponds to the expected size of P1 whereas the RACE PCR product of *ascB* was smaller than the expected size corresponding to the P2 transcript indicating the possibility of new TSS. (C) Figure indicating the predicted RBS of OSS and the duplicated sequence observed in strain ESS. The two putative RBS of strain ESS is shown in boxes. The figure also indicates the mutated RBS

in strain P2E5 obtained through genome engineering. The new transcription start site of *ascB* gene was marked as +1 (D & E) Assay of putative promoter within the *asc* operon. The sequence from +391 from the start codon of *ascF* till the end of the operon was cloned into the *EcoRI* and *KpnI* site of pProbe-NT' vector. *solid lines*, pProbe-A5; *dotted lines*, pProbe-NT'; *solid circle*, MG1655; *solid diamond*, MG1655- ΔP_{ascG}-*ascB*::*frt* and *solid triangle*, MG1655- Δ*chbB-chbF*::*frt*. (D) LB and (E) LB supplemented with 2 g/L glucose.

4.4.4. Optimizing the cellobiose metabolic pathway through oligo-mediated recombineering

In order to further verify if only *ascB* is rate-limiting for cellobiose metabolism or other genes related to *chb* and *asc* operons could enhance cellobiose metabolism, two constitutive promoters (of the *chb* and the *asc* operons) and RBS of 6 genes (*chbB, chbC, chbA, chbF, ascF,* and *ascB*) were randomized through oligo-mediated genome engineering. Consistent with the genotype observed in the strain ESS, efficient cellobiose metabolizing strains obtained through genome engineering had mutations in the upstream region of the *ascB* gene (Fig 4.3A). One representative mutant OSS-P2E5 (shown on Fig 4.3A and Table 4.3) had a growth rate similar to strain, OSS-*ascB**. These results indicate that the *ascB* gene indeed might have a significant role in cellobiose metabolism beyond being present in a minor/incomplete operon.

Strain OSS expresses two different phospho-*β*-glucosidases: ChbF and AscB. However, it was intriguing question that the AscB protein from the *asc* operon (and not ChbF) plays a significant role to enhance cellobiose metabolic ability in strain, ESS and in P2E5; Previous reports indicate that even with high selection stringency on cellobiose, the *asc* operon could not support growth with cellobiose as a sole carbon source [15]. We have previously expressed the *asc* operon under constitutive promoter and the strains could still not grow on cellobiose [18]. The dominant role of *ascB* for cellobiose metabolism (as described in this study) would be a major reason for the conservation of the incomplete operon through the evolution. In accordance with that the *asc* operon orthologs evolved to lose the *ascF* gene but not the *ascB* gene in different lineages of Enterobacter species (Fig 4.5). In addition, the new TSS identified within the *asc* operon indicate that the TSS predicted within the intergenic or coding regions of the gene are also essential regulatory nodes and could serve as a potential target for metabolic engineering and strain optimization. Recent advances in high-throughput screening techniques have reported the presence of such additional Transcription Start Sites (TSS) within an operon [71].



D

```
1     ATTCCGGGGATCCGTCGACCTGCAGTTCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGAAGCAGCTCCAGCCTACACATATACAAGTTTATTCTT 100
                                      +1
101   GACACTAGTCGGCCAAAATGATATAATACCTGAGTACTGTTCACACAGGAAACAGCTATGGCCAAAAATTATGCGGCGCTGGCACGCTCGGTGATAGCGG 200
201   CACTGGGCGGCGTTGATAACATCTCGGCGGTCACGCACTGTATGACGCGGTTGCGCTTTGTTATCAAAGATGATGCACTTATCGACAGCCCGACGTTAAA 300
301   AACCATCCCCGGCGTGCTCGGCGTGGTACGTAGTGACAACCAGTGTCAGGTGATTATCGGCAATACCGTTTCACAAGCCTTTCAGGAAGTCGTCAGCCTG 400
401   CTGCCGGGAGATATGCAGCCCGCACAGCCCGTGGGTAAACCCAAACTCACGCTACGTCGCATTGGTGCGGGGATCCTCGATGCGCTGATCGGCACCATGT 500
501   CACCGCTGATCCCGGCGATTATCGGCGGATCGATGGTCAAACTGCTGGCAATGATCCTCGAGATGAGCGGCGTGCTGACAAAAGGATCGCCGACCTTAAC 600
601   CATTCTGAATGTGATTGGTGACGGTGCTTTCTTCTTCCTGCCGCTGATGGTCGCGGCATCTGCCGCCATCAAATTTAAAACCAATATGTCGCTGGCGATT 700
701   GCCATCGCGGGTGTGCTGGTACATCCGAGTTTTATTGAACTGATGGCGAAAGCGGCCCAGGGTGAACATGTCGAATTTGCCCTGATTCCGGTCACCGCGG 800
801   TGAAATACACCTACACGGTGATCCCGGCGCTGGTGATGACCTGGTGCCTGTCATATATCGAACGCTGGGTGGACAGCATTACACCGGCGGTGACCAAAAA 900
901   CTTCCTCAAACCAATGCTGATTGTGTTGATTGCCGCACCGCTGGCAATCCTGCTGATTGGCCCGATTGGTATCTGGATCGGTAGCGCCATTTCGGCGCTG 1000
1001  GTTTACACCATTCATGGTTATCTGGGCTGGCTTTCAGTCGCCATTATGGGCGCGCTGTGGCCTCTGCTGGTAATGACCGGGATGCACCGCGTCTTTACGC 1100
1101  CAACCATCATTCAGACCATTGCCGAAACCGGCAAAGAAGGGATGGTCATGCCGTCAGAGATCGGCGCTAACCTGTCGCTGGGCGGTTCATCACTGGCAGT 1200
1201  GGCGTGGAAAACGAAAAACCCGGAACTGCGTCAGACGGCGCTGGCTGCGGCGGCATCAGCCATTATGGCGGGGATTTCCGAACCGGCGTTATATGGCGTG 1300
1301  GCGATCCGCCTGAAACGTCCGCTTATCGCCAGTCTTATCAGCGGTTTTATTTGCGGCGCGGTTGCCGGTATGGCGGGGCTTGCCAGCCACTCAATGGCAG 1400
1401  CGCCGGGGCTATTTACCAGCGTGCAGTTCTTCGATCCGGCGAATCCAATGAGCATCGTCTGGGTGTTCGCGGTCATGGCGCTGGCGGTGGTGCTGTCGTT 1500
                                                                                  +1
1501  TATCCTCACACTGTTGCTCGGCTTTGAGGATATTCCTGTTGAGGAAGCGGCTGCCCAGGCGCGAAAGTATCAGAGCGTACAACCGACCGTCGCCAAAGAA 1600
1601  GTAAGTCTTAATTGAGGATGAAAGAGGATGAAAATGTCAGTATTTCCAGAAAGTTTTTTATGGGGCGGCGCGCTTGCCGCCAACCAGTCTGAAGGTGCGT 1700
1701  TCCGTGAAGGTGACAAAGGTCTGACCACTGTCGATATGATCCCACACGGCGAGCATCGAATGGCGGTGAAACTGGGGCTGGAAAAACGTTTTCAGTTGCG 1800
1801  AGATGACGAGTTTTATCCCAGCCATGAGGCGACGGATTTTTATCATCGTTATAAAGAAGATATCGCCCTGATGGCAGAGATGGGATTCAAGGTTTTCCGT 1900
1901  ACCTCAATTGCCTGGAGCCGTCTCTTTCCGCAGGGCGATGAAATCACGCCCAATCAGCAGGGCATTGCTTTTTATCGTTCTGTCTTTGAAGAGTGTAAAA 2000
2001  AGTACGGTATCGAACCGCTGGTCACGTTGTGCCACTTCGATGTGCCGATGCATCTGGTCACCGAATATGGCTCCTGGCGTAACCGCAAGCTGGTGGAGTT 2100
2101  TTTCAGCCGCTACGCCAGAACCTGCTTTGAAGCATTTGATGGTCTGGTGAAATACTGGCTAACCTTCAATGAAATCAACATTATGTTGCATAGCCCGTTC 2200
2201  TCCGGCGCGGGTCTGGTGTTTGAAGAAGGTGAAAATCAGGATCAGGTGAAATATCAGGCCGCGCATCACCAGCTGGTTGCCAGTGCGCTAGCCACCAAAA 2300
2301  TCGCCCATGAGGTTAACCCGCAAAATCAGGTGGGCTGTATGCTGGCGGGCGGTAACTTCTACCCTTACAGTTGCAAGCCGGAAGATGTCTGGGCGGCGCT 2400
2401  GGAGAAAGACCGGGAAAACCTGTTTTTTATCGATGTGCAGGCGCGGGGCACGTATCCGGCTTACTCTGCCCGCGTATTCCGCGAAAAAGGGGTAACCATC 2500
2501  AACAAAGCACCGGGCGATGATGAAATCCTGAAAAACACCGTCGATTTTGTCTCTTTCAGCTATTACGCCTCGCGCTGCGCCTCGGCGGAGATGAACGCCA 2600
2601  ACAACAGCAGTGCGGCGAACGTGGTGAAATCGCTGCGTAATCCGTATCTACAGGTGAGCGACTGGGGCTGGGGAATTGATCCACTCGGTCTGCGTATCAC 2700
2701  CATGAATATGATGTACGATCGTTATCAGAAGCCGCTGTTTCTGGTGGAAAACGGCCTGGGCGCAAAAGATGAATTTGCTGCCAATGGCGAGATTAACGAC 2800
2801  GACTATCGCATCAGCTACTTACGCGAACATATCCGCGCAATGGGCGAAGCGATTGCAGACGGCATTCCGCTGATGGGCTACACCACATGGGGCTGTATTG 2900
2901  ATTTAGTTTCCGCCTCTACGGGTGAAATGAGCAAACGCTACGGCTTTGTCTTTGTTGACCGTGACGACGCAGGCAACGGTACGCTGACGCGCACGCGTAA 3000
3001  GAAATCATTCTGGTGGTATAAAAAAGTGATTGCCAGTAATGGGGAAGATTTAGAG                                              3055
```

Fig. 4.4. Gene specific RT-PCR for *ascB* (A) or ascF (B) in strains OSS, ESS, OSS-*yebK**, OSS-*ascB**. Samples were collected at the indicated cycle number and analyzed on 2% Agarose gel. The *rrsB* (C) gene was used as an

internal control (D) Nucleotide sequence of *ascFB* operon in strain ESS. The scar sequence and the CP12 promoter is indicated in blue; the TSS of *ascF* and *ascB* are indicated in red; the start and stop codons of *ascF* and *ascB* gene are indicated in green and dark blue respectively; the duplicated nucleotide above *ascB* gene in strain ESS is indicated in purple;

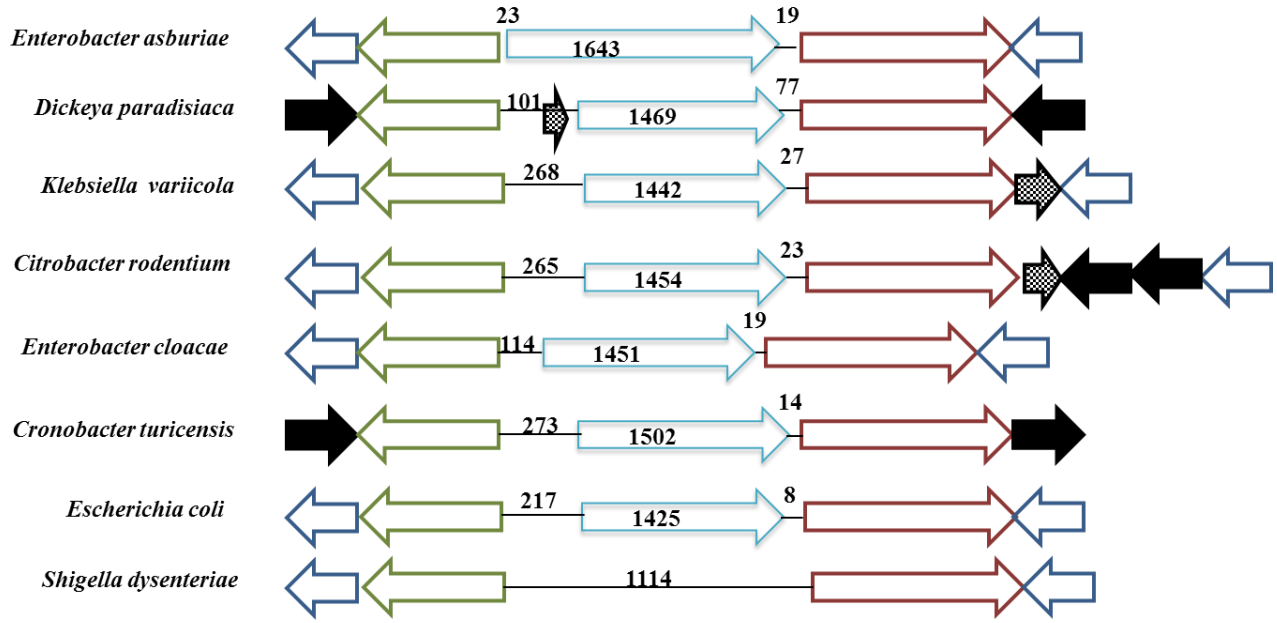


Fig. 4.5. The *asc* operon structure in different *Enterobacter* was analyzed using MicrobesOnline search for homologs. Green arrows indicate *ascG* gene; Blue arrows indicate *ascF* gene; Red arrows indicate *ascB* gene. The misinterrupted genes are in black. The numbers indicate the number of base pairs.

4.4.5. Role of YebK in cellobiose metabolism

The second mutation observed in ESS, that is *yebK* was found to have a dominant role only upon transfer from rich medium to minimal cellobiose medium. Cells pre-cultured on cellobiose minimal medium exhibited negligible impact with respect to the presence or absence of *yebK* mutation indicating its predominance only when transferred from rich medium to minimal medium (Fig 4.7). It could be possible that this transcription factor might probably be regulating the global nutrient down-shift rather than to be specific for cellobiose. In accordance with that the impact of this transcription factor knockout was observed even when cells are transferred from rich medium to glucose-minimal medium (Fig 4.6).

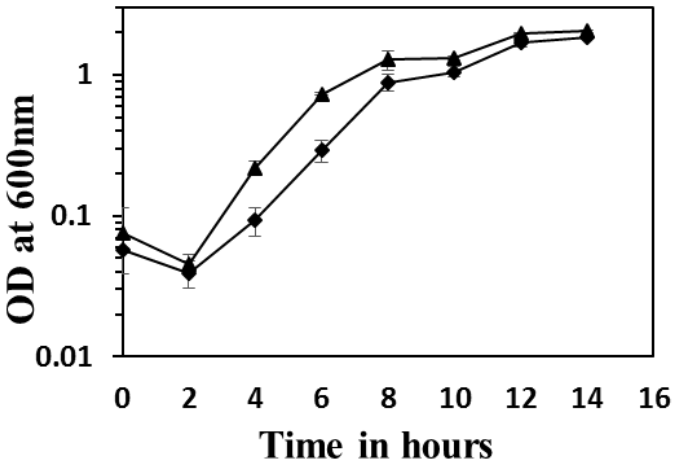


Fig. 4.6. Comparison of growth of *E. coli* on glucose containing minimal medium. MG1655; *closed diamond*, MG1655/Δ*yebK::frt*; *closed triangle*. Error bars indicate the standard deviation of experiments performed in triplicates.

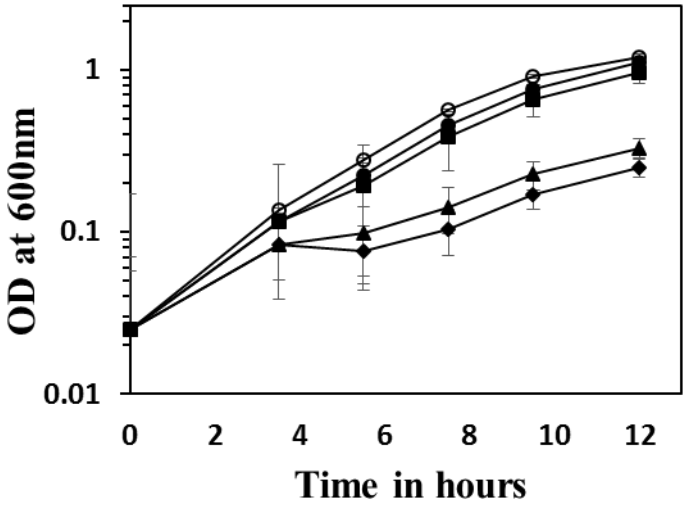


Fig. 4.7. Comparison of growth of *E. coli* strains on cellobiose minimal medium. These strains were pre-adapted on M9-cellobiose minimal medium to mid-log phase and diluted into fresh M9-cellobiose minimal medium. OSS; closed diamond, OSS-*yebK**; closed triangle, OSS-*ascB**; closed square, OSS-*yebK**/*ascB**; closed circle, ESS; open circle.

Mutation in *yebK* helps in reducing the length of lag phase for growth with cellobiose as a sole carbon source

when introduced independently into strains, OSS or OSS-*ascB**. In line with that the specific growth rate on cellobiose in strain, OSS-*ascB** was higher than in strain, OSS (Fig 4.2B). Several factors are proposed to have an influence on the lag phase including the age and size of the initial inoculum, and the physicochemical composition of the new medium i.e. cellobiose [72]. Since all strains used in this study expresses the cellobiose metabolic pathway constitutively the time required to activate cellobiose metabolism may not be the major reason for the lag phase in cellobiose medium. Cell viability was maintained constant for up to 20 hours in LB medium for different strains with and without *yebK* mutation (Fig. 4.8A). Hence, difference in the lag phase upon shift from rich medium to M9- cellobiose minimal medium (between strains with wild type and mutant *yebK*) is not a consequence of difference in cell viability of the initial inoculum. Similarly, viability of strains OSS or OSS-*yebK** was maintained constant throughout the long lag phase indicating that the lag phase is not a due to toxic or osmotic effects of cellobiose (Fig 4.8B.).

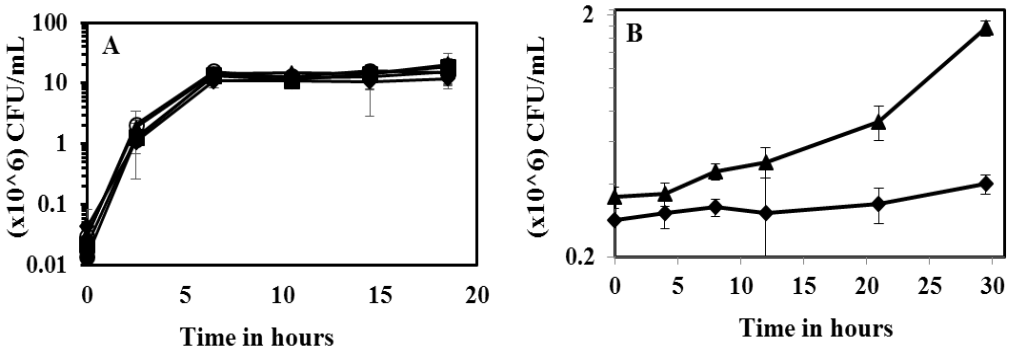


Fig. 4.8. (A) Comparison of growth on LB OSS; closed diamond, OSS-*yebK**; closed triangle, OSS-*ascB**; closed square, OSS-*yebK**/*ascB**; closed circle, ESS; open circle. Samples were collected at the indicated time diluted serially and plated on LB- agar medium. Colonies were counted after 12 hours of plating (B) Comparison of cell viability during the lag phase in OSS (closed diamond) and OSS-*yebK** (closed triangle) growing on cellobiose minimal medium. Samples were collected at the indicated time diluted serially and plated on LB- agar medium. Colonies were counted after 12 hours of plating.

4.4.5.1. yebK recognizes the central metabolic intermediates as co-factor

Previously, it was reported that *yebK* orthologs of *Pseudomonas* and *Shewanella* regulates the ED pathway and gluconeogenesis of the central metabolic pathway respectively, using keto-deoxy-6-phosphogluconic acid (KDPG)

as an effector molecule; it could be possible that *yebK* could also recognize similar effector molecules and regulate the central metabolic pathway. We hypothesized that the lag phase observed in cellobiose minimal medium in strains expressing wild type *yebK* could be because of the requisite to efficiently modulate the central metabolic pathway. Since the target genes regulated by *yebK* is not known, we exploited the auto regulatory role of transcription factors and used *yebK* gene's own promoter in EMSA assay in order to determine if YebK could also recognize the central carbon intermediates like KDPG as a co-factor. As shown in Fig 4.9A, the 6His-YebK binds to its own promoter and the binding is reversed in the presence of KDPG. Interestingly, the truncated YebK (6His-*yebK**) also retained the DNA binding ability (Fig 4.10A).

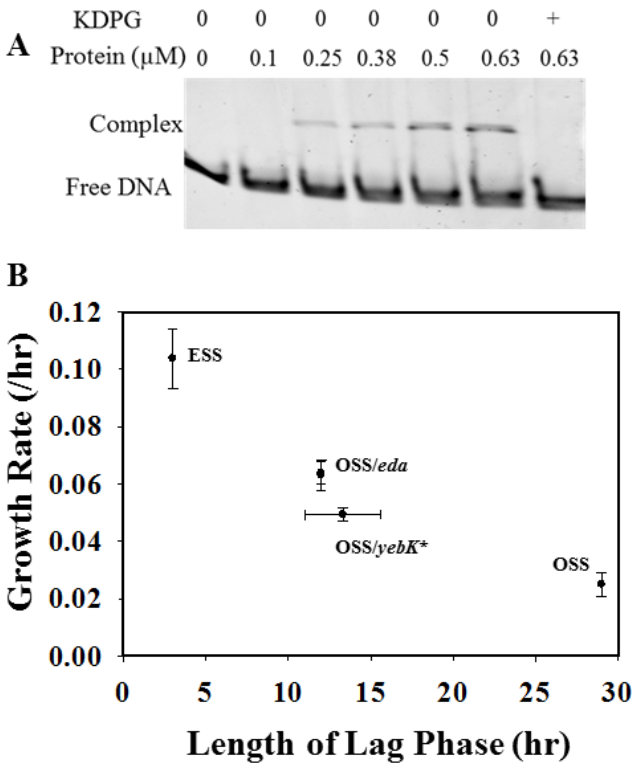


Fig. 4.9. (A) EMSA to analyze the autoregulatory role of 6His-YebK. (B) Comparison of the impact of *yebK** or *yebK* inactivation (through *eda* deletion) on the lag phase for growth on cellobiose. Error bars indicate the standard deviation of experiments performed in triplicates.

4.4.5.2. yebK inactivation is important for controlling lag phase in cellobiose metabolism

In order to further verify if disrupting the DNA binding ability of YebK is essential to counteract the lag phase observed upon transfer from rich medium to cellobiose minimal medium, the *eda* gene encoding the ED pathway

enzyme, KDPG aldolase was deleted in OSS and its growth characteristic on cellobiose was analyzed. Deletion of *eda* gene would result in intracellular accumulation of KDPG which in turn could help in reversing the DNA binding ability of YebK. Similar to that observed in OSS-*yebK**, the lag phase was reduced by at least 16 hours in OSS-Δ*eda* compared to strain OSS (Fig 4.9B). These results provide clear evidence that the lag phase could be reversed by inactivating or impairing the DNA-binding ability of YebK.

Deletion of gene *edd*, encoding the first enzyme of the ED pathway (phosphogluconate dehydratase), results in no production of KDPG and hence the resulting strain OSS-Δ*edd* could not grow on cellobiose minimal medium even after 96 hours of cultivation indicating the need for KDPG to inactivate *yebK* before starting to grow on cellobiose (Fig 4.10B). Expressing *yebK** in OSS-Δ*edd* results in same phenotype as that observed with OSS-*yebK** thus signifying that deregulation of the ED pathway is not the ultimate effect of *yebK* inactivation and there could be other pathways that were controlled by *yebK* using KDPG as one of the signal effectors.

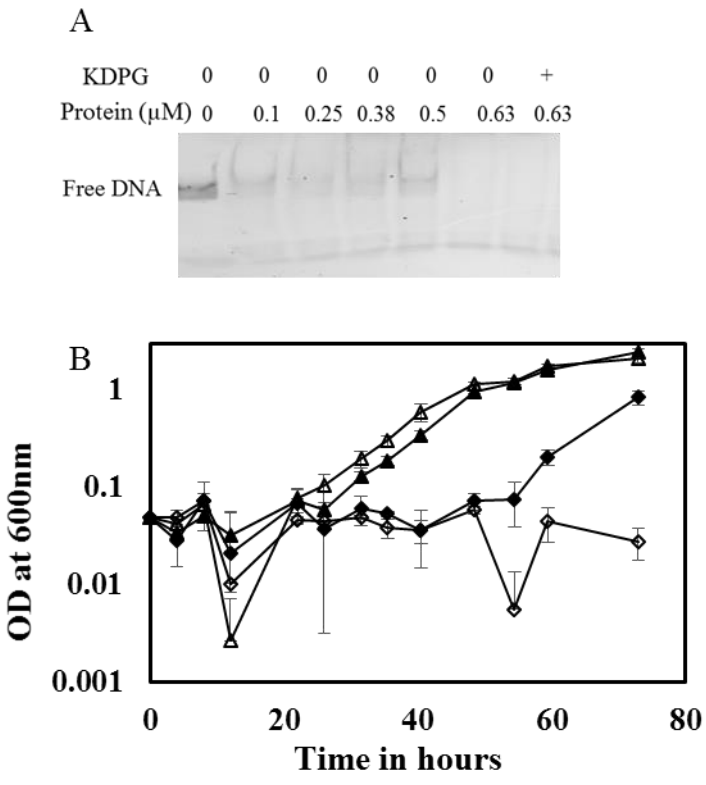


Fig. 4.10. (A) EMSA for 6His-YebK* protein. (B) Growth of *edd* gene deleted strains on cellobiose minimal medium. OSS; closed diamond, OSS-*yebK**; closed triangle, OSS-Δ*edd*; open diamond, OSS-*yebK**/ Δ*edd*; open triangle.

4.4.5.3. Effect of yebK on heterologous metabolic pathway

We then questioned if the mutations in *yebK* could regulate even a heterologous metabolic pathway or is restricted only to the activated cryptic operons of *E. coli*. A heterologous cellobiose metabolic pathway (ECL-00047 to ECL-00051 gene cluster encoding a phospho-β-glucosidase and cellobiose PTS transporter from *Enterobacter cloacae*) was expressed in wild type *E. coli* MG1655 or in MG1655/Δ*yebK* strains. As shown in Fig 4.11., even with heterologous cellobiose metabolic genes, *yebK* mutation could help in reducing the lag phase for growth on cellobiose indicating that *yebK* might act as a global regulation on cellobiose metabolism. Further characterization of genes directly regulated by *yebK* is essential to establish the clear regulatory events controlled by *yebK*.

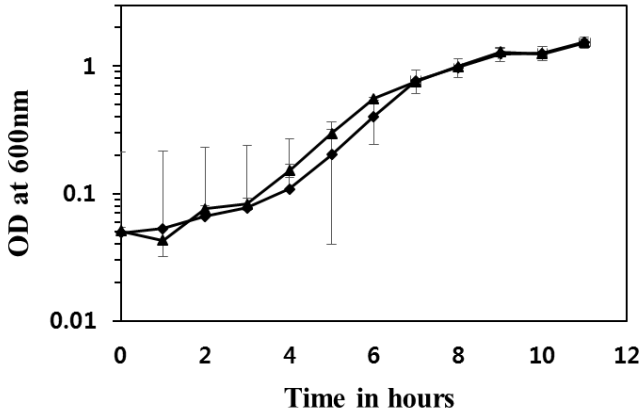


Fig. 4.11. Comparison of growth of *E. coli* expressing pL-cell1 plasmid on cellobiose minimal medium. MG1655; *closed diamond*, MG1655/Δ*yebK::frt*; *closed triangle*. Error bars indicate the standard deviation of experiments performed in triplicates.

While there are several pioneering studies on the transcriptional regulation during stationary phase mediated by *rpoS,* and related genes, it is relatively challenging to study the response of transcription factors controlling the lag phase or transition from one environmental condition to the other. Lag phase is a poorly described phase in bacterial growth stages [72]. Even with *E. coli* (well-studied microorganism) about 40% of the genes are uncharacterized [73] mainly because there is no prior knowledge on the physiological conditions where the gene-of-interest plays a dominant role. Several high throughput screening tools including PM arrays [73], and metabolite profiling [8] were used for the functional assessment of the uncharacterized genes. It is still difficult to characterize the function of putative transcription factors because (except for few regulators) most transcription regulators

would affect the lag phase, specific growth rate or cause flux rerouting without any significant phenotypic changes [74]. However, in this study we demonstrate that the transcription factor *yebK* helps in functionally coupling the minimal nutrient condition to the central carbon metabolism by modulating the length of lag phase relative the specific growth rate of the strain. There could be several speculations on the demand for such modulations in the central carbon metabolism including redox balance, maintenance of the concentration of signaling metabolites, and increasing the energy efficiency. Further studies on the target genes regulated by *yebK* would help in understanding the regulatory changes put forth by *yebK* upon transfer from rich medium to minimal medium.

Several studies were performed in *E. coli* strains adapted on known carbon sources including lactate, acetate, glucose or glycerol[75]. Interestingly adaptive evolution on native (but poorly metabolized) carbon sources (such as lactate or glycerol) resulted in mutations in stress-related regulatory genes (like *rpoS*, *hfq*)[76], global regulators (like *cyaA*, *crp*) [76] or house-keeping genes (like *rpoC*) [75a] whereas adaptation on a non-native carbon source did not have any mutation in such global regulatory genes. Instead, this study provides new insights while engineering *E. coli* for growth on cellobiose though these mechanisms are restricted to the PTS-mediated cellobiose metabolism.

In addition, this study would be the first to report the conditions under which the transcription factor, *yebK* exhibits its impact on *E. coli* growth. The global transcription regulators reported so far regulates a specific nutrient condition. For instance, *crp* is global carbon regulator, *arcA* is regulator of anoxic conditions and *narL* is nitrate/nitrite responsive regulator. Similarly *yebK* could serve as a global regulator controlling the shift in nutrient conditions from rich medium to minimal medium. Further studies are needed to explore in depth the molecular mechanisms of such regulation mediated by *yebK*.

4.5. Conclusions

A combination of recombinant DNA technology (through promoter engineering), adaptive evolution and high-throughput screening helps in addressing the functions of cryptic genes of *E. coli* in greater depths than with the conventional genetics. Through this study we show that the putative TSS identified within the operons through high-throughput technologies like deep RNA sequencing could serve as a potential target for metabolic engineering. In addition, we show the importance of the transcription factor, YebK in global regulation of carbon metabolism upon shift from rich medium to minimal medium.

Chapter 5

Deciphering the Regulatory Changes in strain ESS

(Regulations of the TCA cycle mediated by the transcription factor, YebK)

5.1. Abstract

The glycolytic intermediates play a key role in driving the metabolic flux in bacteria. Thus, if there could be difference in the glycolytic intermediates in strains growing on glucose or cellobiose then despite having a similar growth rate the metabolic end-products obtained would be different from glucose or cellobiose. In line with that, the yield of metabolic end-products like acetate or ethanol is different for glucose or cellobiose in different microbes including yeast and *E. coli*. In bacteria, the transcriptional regulatory network plays a significant role in achieving a functional coupling of the host physiology to the nutrients in the growth medium. Detailed understanding of the endogenous regulatory network is especially important when engineering a heterologous metabolic pathway. In this study, we decipher the interplay of the transcription factor, *yebK* for modulating the central carbon metabolism to be amenable for efficient cellobiose metabolism. Here we show that the TCA cycle activity of glucose and cellobiose are different and that the mutations in yebK could possibly rescue the activity of the TCA cycle.

5.2. Introduction

Studies up-to-date for the construction of recombinant cellobiose metabolizing strains of *E. coli* or yeast have focused on engineering cellobiose transporter and intracellular *β*- glucosidase that converts cellobiose into glucose or glucose phosphate[77]. However, the growth rate of engineered cells on cellobiose is usually lesser than their growth rates on preferred carbon sources, like glucose[77]. Even when the specific growth rate are smilar the yield of the end-products are usually different. This could be because the glycolytic intermediates serve as a driving force to determine the flux direction [78]. Hence, engineering transporter and *β*- glucosidase alone may not be sufficient for efficient cellobiose metabolism. It becomes necessary to understand the cellular responses to cellobiose metabolism and modulate the central metabolic pathway for obtaining maximum energetic benefits from cellobiose in recombinant industrial hosts.

There has been an accumulating wealth of information that the robustness and dynamics of the endogenous metabolic pathway should be coupled to the cellular responses to perturbations in the environmental nutrients[79]. In addition, changes in the concentration of intracellular signaling metabolites (such as pyruvate, phosphoenol pyruvate, cAMP etc) act as a flux sensors to monitor the relative and absolute flux distribution through the central carbon metabolism [78]. Hence, it is always necessary to couple the newly engineered genetic pathway to the endogenous pathway. However, it is difficult to understand the coupling factors especially when engineering the host for the metabolism of cryptic or non-native nutrient.

Often the failure to conquer the complexity of biological systems is because of the gaps in our understanding of the regulatory mechanisms governing the system under investigation. Adaptive laboratory evolution (ALE) has been serving as a basic tool for several decades to understand the metabolic versatility of bacteria under variety of environmental conditions [80] including growth in the presence of organic solvents, high temperature [81] or poor carbon sources like lactate or glycerol [75a, 82]. Here, we extend the molecular characterization of strain ESS and describe the role of novel mutations regulating the cryptic cellobiose metabolism in *E. coli*.

5.3. Materials and methods

5.3.1. Bacterial strains and media composition

All strains and plasmids used in this study are listed in Table 5.1. Bacteria were cultured at 37℃ in Luria Bertani broth (LB) or minimal medium supplemented with cellobiose or the indicated carbon source. Strains carrying temperature-sensitive plasmids were grown at 30℃. Media were supplemented with suitable antibiotics (30 μg chloramphenicol/mL and 100 μg ampicillin /mL). For long-term storage, cells were maintained as 20 % glycerol stocks at –80℃. M9 minimal medium, supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 4 g sugar/L, was used to characterize the cell growth rate of the modified strains. In these tests, overnight cultures grown in LB were collected, washed once with M9-salts and suspended to a final OD of 0.05 in 50 mL of M9 medium supplemented with the test sugar in a 250-mL flask, and the cultures were incubated at 37℃ with rotation at 200 rpm. Cell growth was monitored by measuring the optical density at 600 nm (OD_{600}) every 3 hours with a Biochrom Libra S22 spectrophotometer.

Table 5.1. *E. coli* strains and plasmids used in this study

Strains/plasmids	Description/genotype	Reference/source
Strains		
E. coli MG1655	Wild type	[37]
MG1655/Δ*yebK*::*frt*	MG1655 with *yebK* gene deleted	This study
OSS	MG1655 with CP12chb plus CP12asc	[18]
ESS	OSS adapted in cellobiose for 30 days	[18]
OSS-Δ*aceA*:*frt*	OSS with *aceA* gene deleted	This study
OSS-Δ*kgtP*:*frt*	OSS with *kgtP* gene deleted	This study
Plasmids		
pSIM5	λ-Red recombinase expression plasmid and temperature-sensitive replication	[62]
pCP20	Yeast FLP recombinase gene controlled by cI repressor and temperaturesensitive replication.	[39]
pKD13	Template plasmid for gene disruption. The kanamycin resistance gene is flanked by FRT sites. oriR6K-gamma origin requiring the pir+ *E. coli*.	[63]
pET31b+*yebK*-6his	pET31b+ plasmid with *yebK*-6his	This study

pET31b+*yebK*-flag	pET31b+ plasmid with *yebK*-flag	This study
pZB-K	pZB plasmid expressing *kgtP-rfp* fusion protein from P_{zt1}	This study
pZB-KY	pZB plasmid expressing *kgtP-rfp* fusion protein from P_{zt1} and *yebK* from P_{BAD}	This study

5.3.2. Strain construction

Gene deletion was performed using the λ-Red recombination system, as described previously [63]. Briefly, the kanamycin cassette of pKD13 was amplified using the deletion primers containing 50bp homology to the target genes. Cells carrying the λ -Red system, under the control of the P_L promoter (pSIM5), were induced at 42℃ for 15 minutes, made electro competent, and then transformed with the PCR product. Transformant colonies carrying the desired modification were directly selected on LB agar plates supplemented with kanamycin. The kanamycin cassette was then cured using the FLP recombinase expressed from pCP20 plasmid. Genomic DNA was isolated from the transformants, and the target region was PCR amplified and sequenced to confirm site-specific insertion.

5.3.3. ChIP-sequencing

For ChIP-seq experiments, YebK with *flag* tag was expressed from plasmid. The pET31b-*yebK-flag* plasmid was then transformed into BL21 (DE3) strains and grown to OD 2.0 in 50 mL LB medium at 37℃. Cells were then induced with 0.5 mM IPTG and grown at 30℃ for 3 hours. Cells were cross-linked with formaldehyde for 10 minutes and the reaction terminated by the addition of excess glycine. Cells were collected, washed with 1X PBS and suspended in lysis buffer supplemented with lysozyme and protease inhibitor. Cells were sonicated, diluted in IP dilution buffer and non-specific proteins adsorbed on protein A agarose bead. The supernatant was then subjected to immunoprecipitation with specific antibody (IgG antibody or flag antibody) and adsorbed on protein A bead. The eluted DNA was heated at 62℃ for 4 hours to reverse the cross-linking and the DNA purified by Qiagen kit and sequenced using illumina sequencer.

5.3.4. Protein purification

E. coli BL21(DE3) transformants with pET31b+yebK his tagged or pET31b+yebK* his tagged plasmid was grown in 250 mL LB broth to OD_{600} of 0.7, and protein expression was induced with 0.5 mM IPTG and grown at 18℃ overnight. Cells were harvested, suspended in lysis buffer (50 mM Tris-HCl pH7.8, 1 mM DTT, 300 mM

NaCl, 10 mM imidazole supplemented with RNase A) with or without 10 mM alpha -ketoglutarate (pH 8.0) and then sonicated. Cellular extracts were then purified using Nickel-NTA agarose column. The proteins were eluted in elution buffer (50 mM Tris-HCL pH 7.8, 1 mM DTT) with gradient concentration of imidazole (10-500 mM). For purification of YebK, 25 mM alpha-ketoglutarate was supplemented in the elution buffer. Purity of the proteins was analyzed on 10 % denaturing acrylamide gel. The oligomeric state of the protein was analyzed using analytical gel filtration chromatography on an Akta fast protein liquid chromatography system.

5.3.5. Electrophoretic Mobility Shift Assay (EMSA)

The template for EMSA was PCR amplified from MG1655 genomic DNA. 150 ng of the DNA was incubated with proteins in the concentration ranging from 0 to 1 µM in binding buffer (16 mM Tris pH7.5, 3 mM $MgCl_2$, 30 mM NaCl, 0.0065 % Triton-X and 0.033 mg/mL BSA) for 1 hour at 37° C. When indicated, KDPG was substituted. The reaction was analyzed on 3 % agarose gel.

5.3.6. Thermal shift assay

Purified 6His-yebK was appropriately diluted in a buffer containing 20 mM Tris-Hcl, 500 mM NaCl, 1 mM DTT. All assays were performed using 5 µM protein per well in 25 µL volume containing 20X Sypro Orange dye. When indicated alpha-ketoglutarate or KDPG was added. 96-well PCR plates were sealed with optical seal and thermal scanning (25℃ to 95℃ at 1℃/min) was performed using a Roche Light Cycler 480. All experiments were performed in triplicate. The pre-melt and post-melt fluorescence signals of each sample are normalized to 0 % and 100 % respectively.

5.4. Results and discussion

5.4.1. Impact of changes in central carbon metabolism on cellobiose metabolism

The transcription factor, *yebK* is identified as one of the two beneficial mutations of strain ESS that was adapted on cellobiose minimal medium. YebK belongs to the RpiR family of transcription factor and is known to regulate the central carbon metabolism in different bacteria including *Pseudomonas* and *Shewanella* [83]. In *E. coli* also, the transcription factor *yebK* was verified to be involved in the global regulation of several different carbon sources (Fig 5.1). The impact of this transcription factor was particularly predominant when transferred from rich medium to minimal medium and the effect was positive on few carbon source (including glucose, glycerol, and alpha-

ketoglutarate) and negative on few other carbon sources (including acetate, fatty acid).

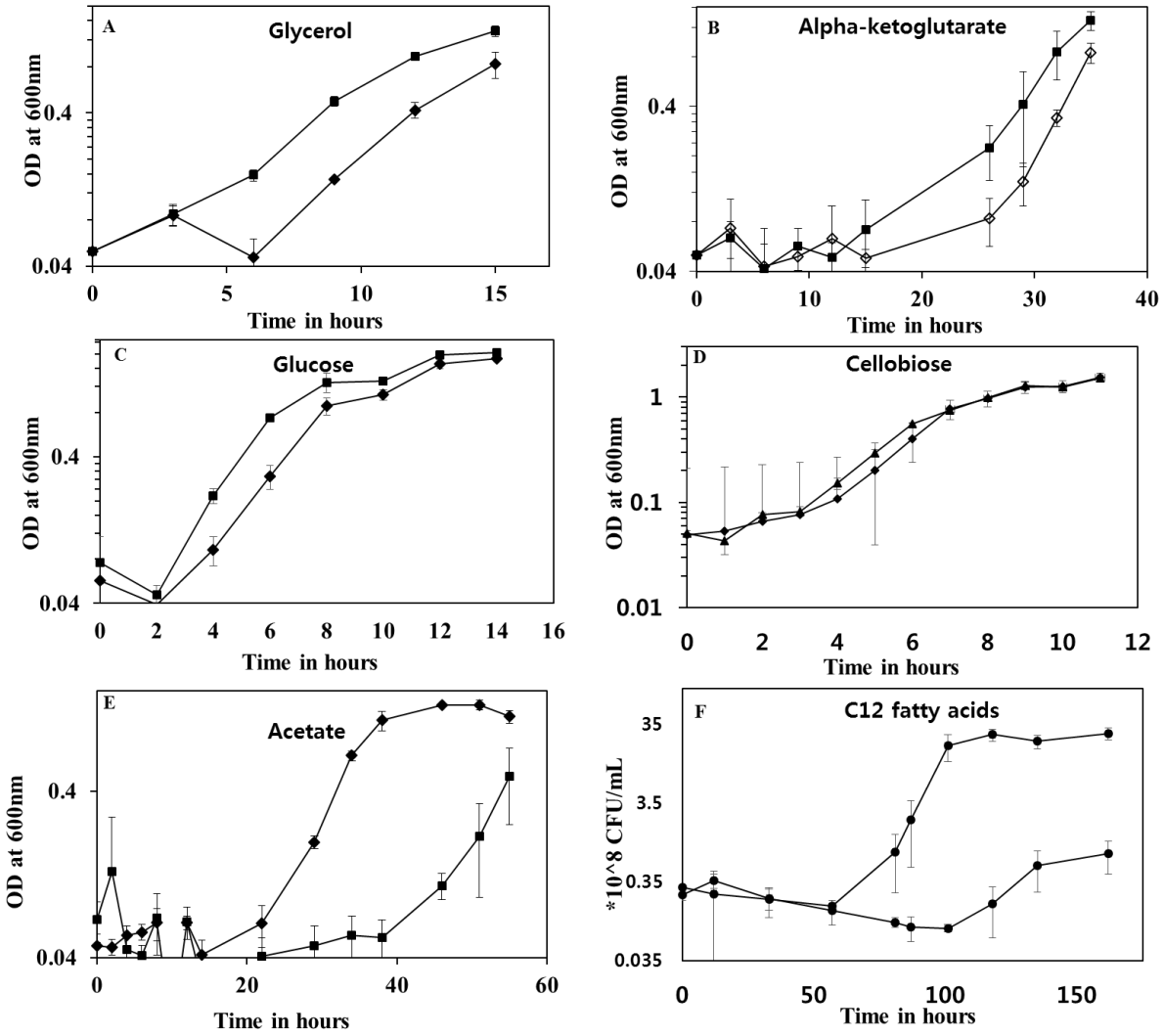


Fig 5.1. Growth on minimal-medium with different carbon sources (A) Glycerol; (B) Alpha-ketoglutarate; (C) Glucose; (D) Cellobiose (heterologous); (E) Acetate; (F) C12 fatty acid. MG1655, *closed diamonds*; MG1655/Δ*yebK*::frt, *closed square*. For these experiment cells were grown overnight in LB, washed with M9 salts and re-suspended in M9-minimal medium with the suitable carbon source. Summary of this data is described in Chapter 5.

As described in chapter 4, it is interesting that such a global transcriptional regulator exhibits a positive impact on cellobiose metabolism. Hence, we hypothesized if changes in central carbon metabolism could help in enhancing the cellobiose metabolic rate. The pathways down-regulated in fast cellobiose metabolizing strain ESS compared

to the slow cellobiose metabolizing strain was previously reported through transcriptomic analysis. The strain OSS grown on cellobiose exhibit a higher expression level of the glyoxylate pathway (both at the mRNA and the protein level) when compared to strains grown on glucose (Fig 5.2). The glyoxylate pathway functions to recycle carbon without a net yield of reducing powers such as NADH. The deletion of the glyoxylate shunt pathway helps in enhancing the cellobiose metabolic rate in strain OSS indicating that the changes in central carbon metabolism (particularly, TCA cycle) can help in enhancing the cellobiose metabolic rate (Fig 5.2).

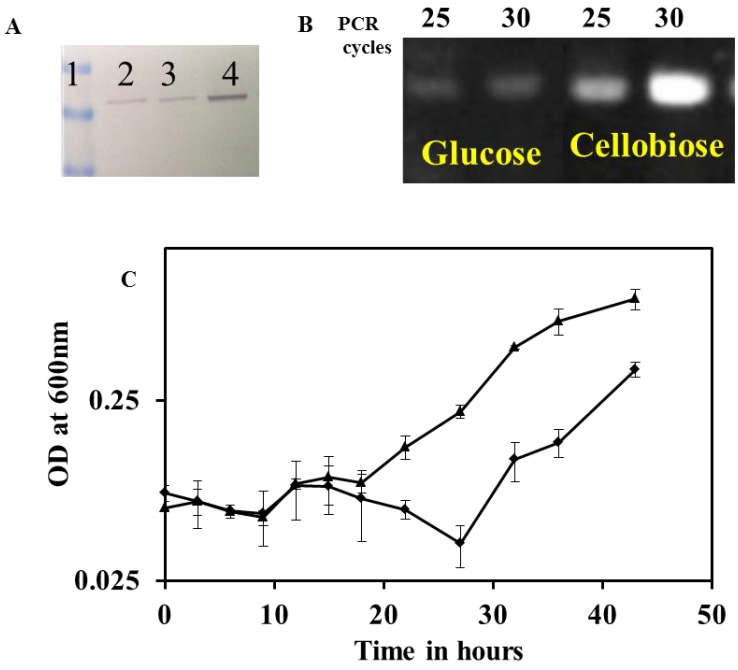


Fig 5.2. (A) Western blot for strain OSS::*aceB-flag* grown on different carbon source to an OD of 0.5. Lane 1, *protein ladder*; Lane 2, *aceB-flag* expressed in strains grown on *glucose minimal-medium*; Lane 3, *aceB-flag* expressed in strains grown on *xylose minimal-medium*; Lane 4, *aceB-flag* expressed in strains grown on *cellobiose minimal-medium*. (B) Semi-quantitative RT-PCR for *aceB* transcript for strain OSS grown on glucose or cellobiose-minimal medium. (C) Growth on cellobiose of strain OSS (closed diamond) or strain OSS-Δ*aceA*:*frt*.

5.4.2. Identifying genes regulated by *yebK*

Since *yebK* is an uncharacterized transcription factor in *E. coli*, ChIP-seq was performed to identify the target genes regulated by YebK. Flag-tagged-YebK was over-expressed from plasmid, cross-linked with formaldehyde and the putative target genes regulated by *yebK* were identified through ChIP-seq. 19 putative sequences identified

from ChIP-seq was assembled using MEME motif prediction tool and consensus YebK binding site was identified (Fig 5.3A and Table 5.2). The derived consensus sequence (tTGTAAnwwwnTTACAa) was more similar to the reported consensus for RpiR family of transcription factor (tTGTAATwwwATTACa) [12]. This ChiP-seq approach though could not identify all possible target genes regulated by *yebK*. Many of the target genes identified by ChIP-seq are uncharacterized or genes of unknown function. Of the genes of known functions at least 3 genes were related to the TCA cycle of *E. coli* (including *pdhR*, *cyoC* and *kgtP*) whereas few others are sugar specific transporter (including *xylF* and *lacY*). Since several carbon sources with positive or negative effect upon yebK deletion are associated with TCA cycle for its metabolism (including acetate, fatty acids and cellobiose) and several genes of TCA cycle are identified through ChIP-seq, we hypothesized that controlling the TCA cycle activity would be one of the possible regulations modulated by *yebK*.

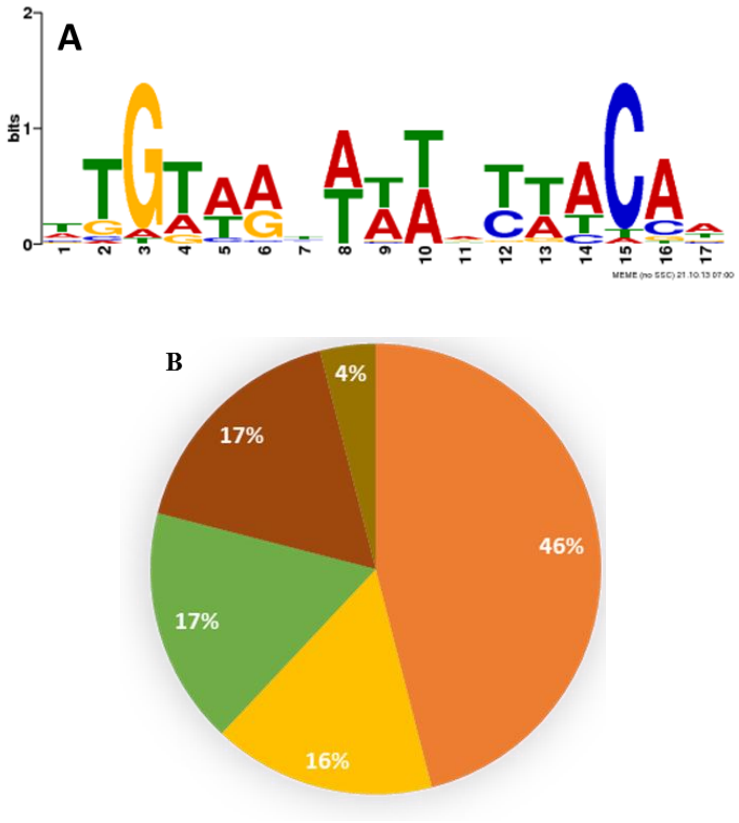


Fig 5.3. (A) YebK consensus sequence obtained from MEME upon assembling the putative target sequences enriched in ChIP-seq. (B) Distribution of YebK binding site relative to the coding sequence or the promoter. *Red*, percentage of genes with YebK binding site within the promoter region or in the intergenic region between a promoter and the coding sequence. *Orange*, percentage of genes with YebK binding site above the promoter of a

gene but present within the coding region of an upstream gene. *Green*, percentage of genes with YebK binding site within an operon. *Brick red*, percentage of genes with YebK binding site in the coding region of a gene. *Brown*, percentage of genes with YebK binding site downstream of a genes.

5.4.3. *In vitro* analysis of YebK binding

5.4.3.1. YebK purification and analysis of its oligomeric state

The purification of active YebK protein was challenging and the proteins obtained were in its higher oligomeric state when analyzed through gel filtration chromatography. Hence, several TCA cycle intermediates (around the putative target genes regulated by yebK) where included in the purification buffer (including alpha-ketoglutarate, succinate, L-glutamine and pyruvate) and the oligomeric state of the purified protein was analyzed by gel filtration chromatography. Interestingly, the protein isolated in the presence of alpha-ketoglutarate retained as a trimer whereas in the presence of other intermediates the protein was still in its oligomeric state (Fig 5.4A). Hence, we hypothesized that alpha-ketoglutarate could serve as a prosethetic group of apo-YebK. Fluorescence-based thermal shift assay was performed to further verify the stability of the proteins purified in the presence or absence of alpha- ketoglutarate. The denaturation kinetics indicate that the protein purified in the presence of alpha-ketoglutarate was more thermo stable (Fig 5.4B). Since KDPG was previously reported to be a potential effector of HexR family of protein, we also compared the efficiency of KDPG or alpha-ketoglutarate to enhance the thermo stability of apo-YebK (Fig 5.4C and 5.4D). According to the denaturation kinetics of the protein, it could be possible that both alpha-ketoglutarate and KDPG could serve as a co-factor of yebK while alpha-ketoglutarate shows efficient binding at a lower concentration and KDPG shows efficient binding at a higher concentration.

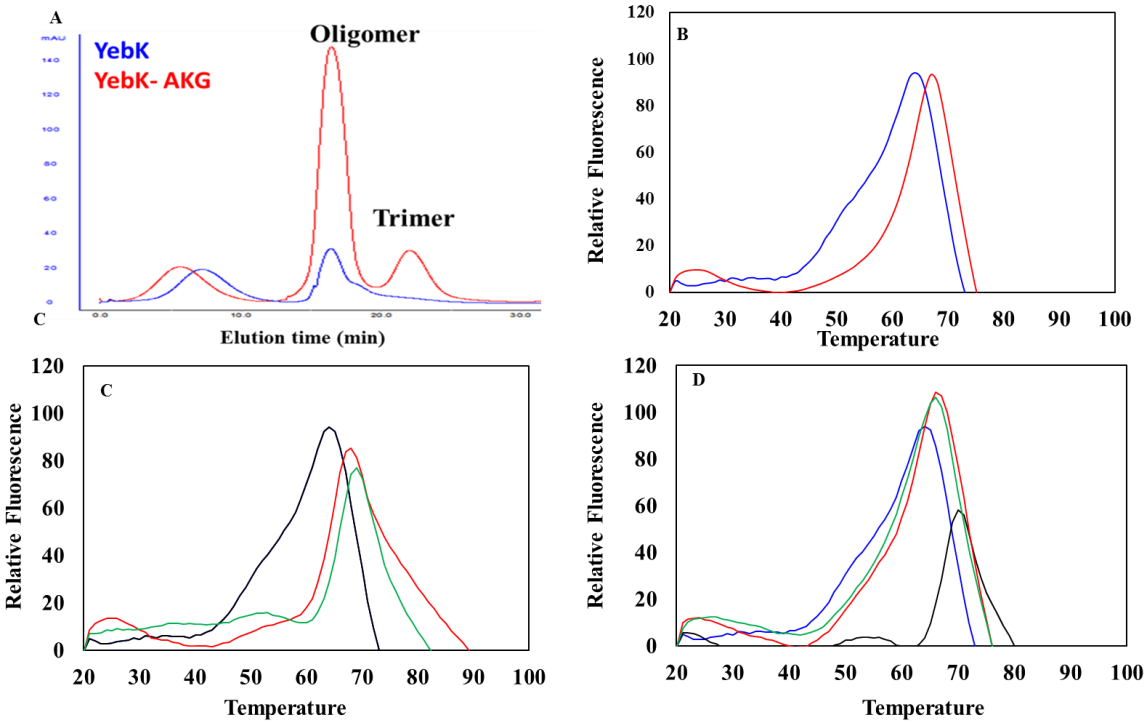


Fig. 5.4. (A) Analysis of the oligomeric state of apo-YebK and YebK purified in the presence of alpha-ketoglutarate. (B) Fluorescence based thermal shift assay for apo-YebK (blue) or YebK purified in the presence of alpha-ketoglutarate (Here then referred to as YebK-AKG) (red). Data represents an average of three independent replicates. Fluorescence based thermal shift assay for apo-YebK in the presence of KDPG (C) or alpha- ketoglutarate (D). Apo-YebK, *blue line*; 0.5 mM co-factor, *black line*; 1.5 mM co-factor, *red line*; 2 mM co-factor, *green line*. Data represents an average of three independent replicates.

Table 5.3. List of target genes identified through ChIP-seq.

Gene name	Fragment size	Coverage		YebK binding site	Gene Function	Location relative to the start codon
		Control	Sample			
yibI	101	34	382	GTGTAAAAAAGCAACAA	Unknown Protein	-308bp
xylF	101	40	372	CTGTAATTGTTTTCCCC	Xylose transport	-96bp
yihU	151	32	476	TTGACGATAAATTACAA	Sulfoquinovose metabolism	-83bp
yrdA	101	28	318	TATTAACTTTACTCCCT	Carbonic anhydrase	-719bp

yjfI	101	23	290	ACGTAATTATCTTACCA	Conserved protein	-66bp[ψ]
ybbK	101	25	412	ATGTTTATTACTTTCAT	Predicted protease	-661bp
ybjL	101	18	243	TTATCGTAAATTACCAT	Predicted transporter	-48bp
pdhR	101	34	426	AGTTTGAATTGTTACAA	Pyruvate dehydrogenase	-434bp
rrsD	151	19	401	GGGTGAAAAAACAACAA	16S rRNA	-353bp
yedW	101	17	338	AGGGAAGAAACTTTCAA	Transcriptional regulator of two component system for drug resistance	-329bp
gcvA	151	17	293	AGGTAGCTTTGCTACCA	Transcriptional regulator of glycine cleavage enzyme system	-251bp
tsx	151	30	552	GTAAAAGTTTGCAACAA	Nucleoside transporter	-211bp
yibN	151	28	525	GCGTAATTTTACTACAG	Sulfur transferase	-208bp
yebK	201	22	377	ATGAAAAAAAATAACAA	Unknown transcription factor	-123bp
insG	151	43	383	CTGATAGTTTGTTACAA	Transposase	-111bp
yncC	101	13	338	CTGAAATTATGCAACAA	Transcriptional regulator for biofilm formation	+677bp
kgtP	101	16	298	CGAAAGGAATACGCCAT	Alpha ketoglutarate transporter	+589bp[ψ]
lacY	151	33	419	CTGTTACAATACAACAT	Lactose uptake	+291bp[£]
polA	151	21	402	CTGAAAAAAAATTACAA	Polymerase I	+2890bp[€]
cyoC	101	0	383	ATGTACAAAAACAACAA	Cytochrome *bo* terminal oxidase subunit III	+275bp[£]
waaO	151	24	573	TCGTTGTTATGTAACAA	α- 1,3-glucosyltransferase	+157bp[ψ]
yieN	151	34	822	GTGAAATAAAACAACAA	Regulatory ATPase (kanamycin resistance)	+1330bp
rplV	101	12	237	TTGACCTACACCAACAA	50S ribosomal subunit	+112bp[£]
yqjA	151	29	442	AAGGAACAAAATTACAA	L-tartrate dehydratase	+106bp[ψ]

[ψ], binding site is oriented in the opposite to the direction of the gene.

[€], binding site present in the intergenic region between the ends of two genes.

[£], the yebK binding site is in the gene present within an operon

5.4.3.1. EMSA

The sequence enriched in the ChIP-seq was used as a DNA probe in EMSA assay. The efficiency of apo-YebK or YebK-AKG to bind to P_{yebK} promoter (auto-regulatory) was verified through EMSA. As indicated YebK-AKG binds the target promoter more efficiently than the apo-YebK and the binding is reversed in the presence of KDPG (Fig 5.5A and 5.5B). According to the binding stoichiometry, it could be possible that YebK protein binds the DNA as a hexamer. Other putative targets of the TCA cycle, *cyoC*, *pdhR*, and *kgtP* were then analyzed for its efficiency to bind apo-YebK or YebK-AKG (Fig 5.6). Similar to that observed with P_{yebK} promoter, YebK-AKG

binds the DNA more efficiently than apo-YebK.

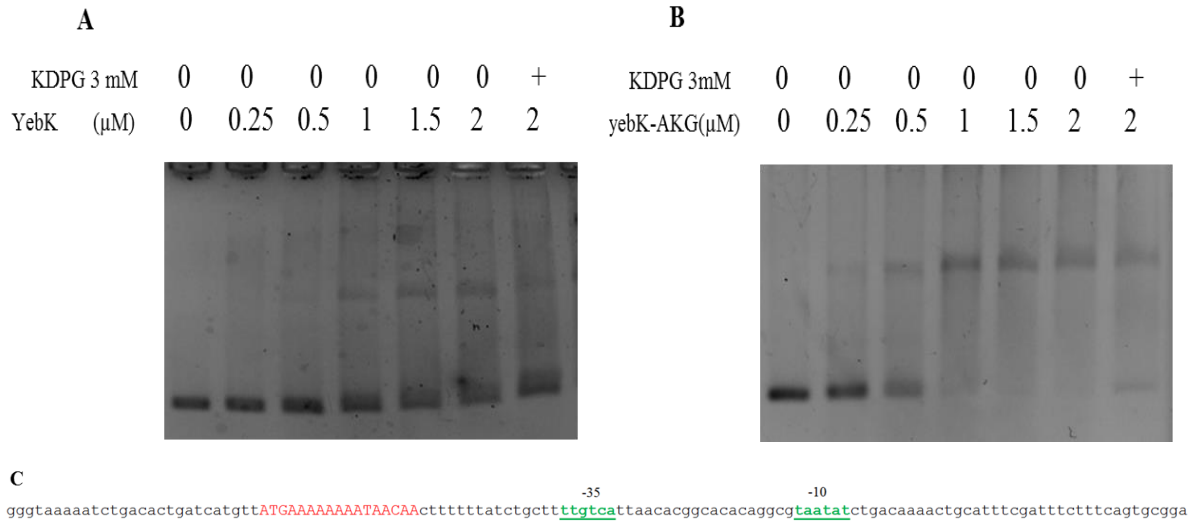


Fig 5.5. Electrophoretic mobility shift assay for the promoter region of P_{yebK} by pruified apo-YebK (A) or YebK-AKG (B). The sequence of the DNA probe used in EMSA assay is indicated in (C). The YebK binding sequence is marked in *red* and the -35, -10 sequences of the promoter is marked *green*. The concentration of protein used (in µM) is indicated in above each lane.

The *pdhR* regulon of *E. coli* regulates the expression of pyruvate dehydrogenase and succinate dehydrogenase complexes (two key enzymes of the TCA cycle) and is regulated by several different transcription factors (Fig. 5.7). Here, we show that (in addition to *pdhR*, *crp*, *fnr*) the *pdhR* regulon is also regulated by *yebK*. It could be possible that YebK performs a global regulation by modulating the signal from the glycolytic pathway (through KDPG) and TCA cycle (through alpha- ketoglutarate) and regulate the expression of *pdhR* regulon. It is also interesting to note that the *cyoC* gene within the *cyo* operon also has a *yebK* binding site. Though the physiological significance of such binding sites is not known, there are increased reports on internal promoters or transcription factor binding sites within an operon.

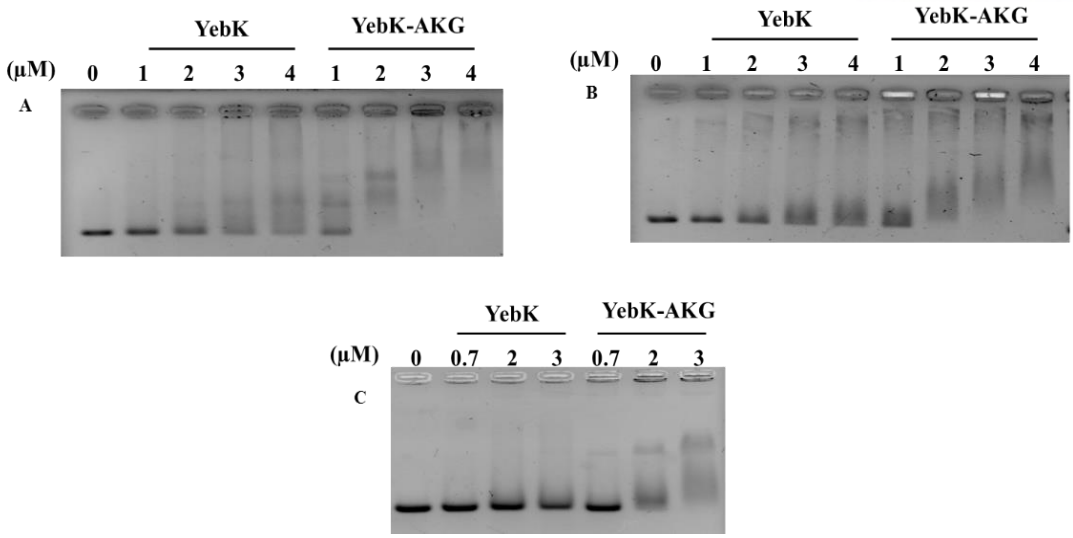


Fig 5.6. Electrophoretic mobility shift assay for (A) *cyoC* gene (B) *pdhR* gene (C) *kgtP* gene . The DNA probe corresponds to the sequence enriched in ChIP-seq.

```
                Start codon of AroP
1   catgaaacctcgtgcggtggttgtttttttgatctacgcagtgatgcgtg
                    TyrR                                Crp-6
51  tgtaagtttgcaattccgttgttgtattaatttgtttacatcaaagaag
              YEBK
101 tttgaattgttacaaaaagacttccgtcagatcaagaataatggtatgcg
                -10 AroP                -35-AroP
151 gcagcgaatgcacccgctttatgcatggttgaagatgagttgcttaaaaa
201 gaaaccgtttgtaaagctcagcctcaacccctctcaatatgtagaatgaa
251 tttaaattcgttttaattgaattaaaaatcaaaaaattggtaagtgaatc
        cra                     crp5
301 ggttcaattcggatttttatagtttaataatcgttaaaaaactcctttcc
                        crp4
351 tacgtaaagtctacatttgtgcatagttacaactttgaaacgttatatat
          crp3          Crp 2           fnr
401 gtcaagttgttaaaatgtgcacagtttcatgatttcaatcaaaacctgta
     -35- pdhR                  -10-pdhR            pdhR
451 tggacataaggtgaatactttgttactttagcgtcAcagacatgaattg
                        Crp 1
501 gtaagaccaatttgacttcggcaagtggcttaagacaggaactcatg
                                        Start codon of PdhR
```

Fig. 5.7. Nucleotide sequence of the P_{pdhR} of *E. coli*. The region used as a DNA probe in EMSA assay is underlined.

5.4.3. Role of YebK target genes on cellobiose metabolism

As shown in Fig 5.2., perturbations in the TCA cycle leads to an increase in the growth on cellobiose. We then hypothesized if any of the putative target genes identified in this study could influence the cellobiose metabolism similar to strain, OSS-*yebK**. As indicated in Table 5.3., several target genes of *yebK* encode for protein of unknown function. In addition, many of the putative targets of *yebK* are present within the coding region of the gene. Though there are increased reports on such binding sites within the intergenic and coding regions, its importance in genetic regulation is not established yet [84]. Hence, the gene kgtP was used to verify the effect of the *yebK* binding sites present within the coding sequence of the gene. The gene, *kgtP*, encodes for an alpha-ketoglutarate transporter and the *yebK* binding site was identified with the coding region of this gene. The gene YebK is expected to bind to *kgtP* and lead to premature termination of transcription of this gene. In theory, presence of YebK* would help in permanently repressing *kgtP* expression. Consistent with that over-expression of *kgtP* gene from a plasmid completely inhibits growth on cellobiose and deletion of *kgtP* helped in enhancing the cellobiose metabolism (Fig 5.8). The growth rate of OSS-K was close to the growth rate of OSS-*ascB** and higher than OSS-*yebK**. These results might indicate two important genetic events: (1) cellobiose metabolism could be enhanced by the changes in the central carbon metabolism itself (2) the putative binding sites present within the coding regions of the gene could also serve as an important target for regulations.

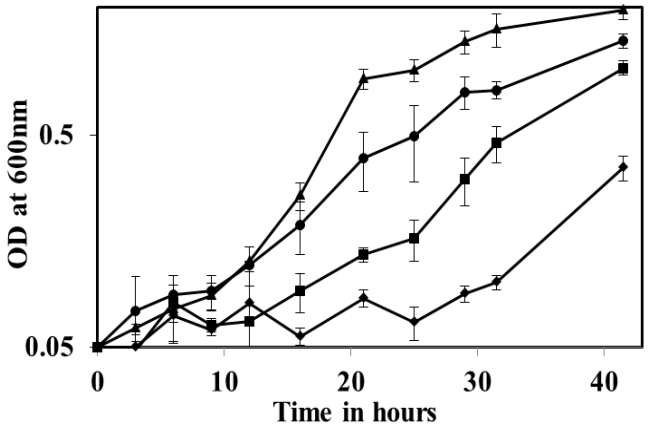


Fig. 5.8. Growth of different strains on cellobiose-minimal medium. OSS, *closed diamond*; OSS-*yebK**, *closed square*; OSS-*ascB**, *closed triangle*; OSS/Δ*kgtP*::frt, *closed circle*.

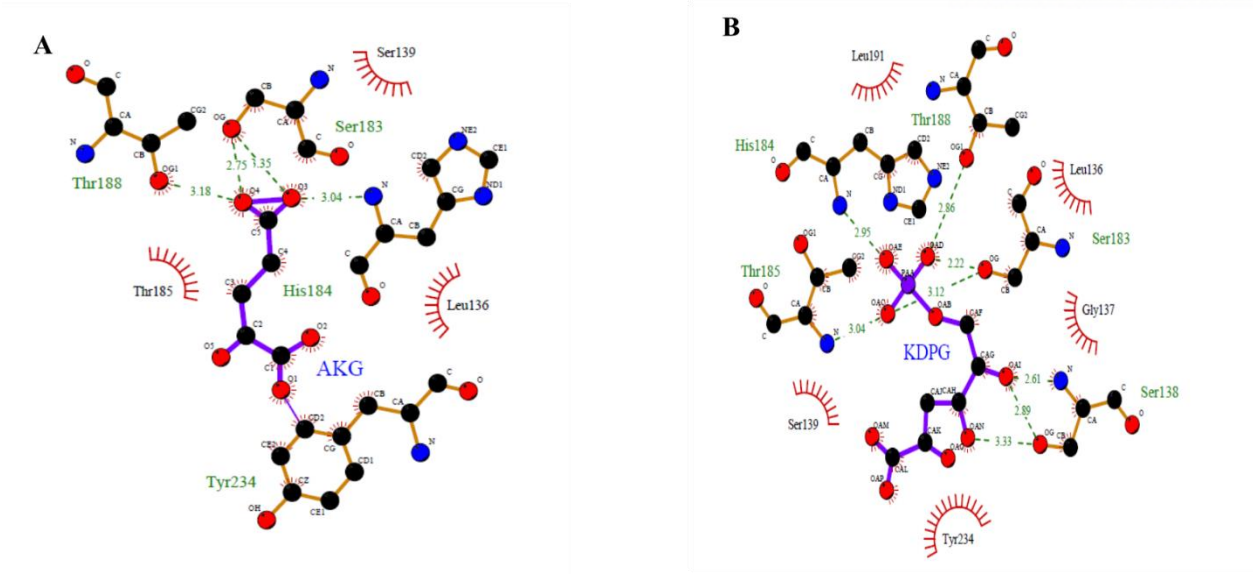


Fig. 5.9. Homology based protein model of YebK bound to alpha-ketoglutarate (A) or KDPG (B). The key residues interacting with the oxygen atoms of alpha-ketoglutarte are Tyr 234 (O_1), Ser 183 (O_3, O_4), Thr 188 (O_4). The key residues interacting with the oxygen atoms of KDPG are Ser 183 (O_{AC}), Thr 185 (O_{AC}), Thr 188 (O_{AD}), His 184 (O_{AE}), and Ser 138 (O_{AN}, O_{AI}). Model was developed by aligning the residues of YebK to the known structure of a similar protein, NanR. (Percentage similarity between yebK and NanR is listed in Appendix IV) The model was credited Dr. Sang Woo Kim of Prof. Sung Kuk Lee's lab.

5.5. Conclusions

Through this study we demonstrate that the transcription factor, *yebK* acts as a global regulator influencing growth on different carbon sources including cellobiose predominantly through changes in the TCA cycle activity. We show that the transcription factor YebK is active in the presence of central carbon intermediates, alpha-ketoglurate and KDPG. Further analysis is required to monitor the interaction between the co-factors with YebK or the physiological significance of such interactions. Though efficient cellobiose metabolism could be achieved through mere over-expression of the transporter and the enzyme, the difference in the metabolic pathway could lead to a reduced yield of the metabolic end-products. It becomes necessary to establish a coupling between the host physiology and the carbon metabolism. YebK serves as protein that couples a carbon source to the central metabolic pathway. Further studies to reveal the entire regulon of YebK would help in understanding the molecular features of YebK.

Chapter 6

Industrial applications of strain ESS

Application I: Co-metabolism of cellobiose and xylose

(ESS as a potential platform strain for SSF, SHF and CBP process)

6.1. Abstract

Natural ability to ferment the major sugars (glucose and xylose) of plant biomass is an advantageous feature of *Escherichia coli* in biofuel production. However, excess glucose completely inhibits xylose utilization in *E. coli* and decreases yield and productivity of the major sugar fermentation due to sequential utilization of xylose after glucose. As an approach to overcome this drawback, *E. coli* MG1655 was engineered for simultaneous glucose (in the form of cellobiose) and xylose utilization by a combination of genetic and evolutionary engineering strategies. The recombinant *E. coli* was capable of utilizing 6 g/L of cellobiose and 2 g/L of xylose in approximately 36 hours, whereas wild-type *E. coli* was unable to utilize xylose completely in the presence of 6 g/L of glucose even after 75 hours. The engineered strain also co-utilized cellobiose with mannose or galactose; however, it was unable to metabolize cellobiose in the presence of arabinose and glucose. Successful cellobiose and xylose co-fermentation is a vital step for simultaneous saccharification and co-fermentation process and a promising step towards consolidated bioprocessing.

6.2. Introduction

Lignocellulosic biomass-derived fuels are considered an efficient alternative to fossil fuels. However, the major hindrance in the use of lignocellulosic biomass is its heterogeneous nature that yields a mixture of sugars, particularly rich in glucose and xylose [30]. Therefore, complete and rapid assimilation of these sugars is a major determinant of the cost, efficiency, and ease of cellulosic fuel production[19, 85]. *Escherichia coli* has a natural ability to utilize both glucose and xylose, and is thus one of the efficient workhorses for cellulosic fuel production[86]. However, carbon catabolite repression (CCR) that leads to a preference for glucose limits the potential of *E. coli* as an efficient biocatalyst for cellulosic fuel production. Until the complete depletion of glucose, *E. coli* cannot metabolize xylose. In mixed-sugar fermentation, xylose metabolism is completely inhibited if glucose concentration exceeds 40% of the total sugar concentration [87]. This complex regulatory system that favors preferential utilization of sugars impedes the fermentation process, leading to decreased yield and altered downstream processing due to the accumulation of unutilized carbon sources [86a].

Catabolite derepressed strains of *E. coli* have been used for simultaneous utilization of major sugars of the hydrolysate (glucose, xylose, and arabinose) [88]. Several *E. coli* strains carrying mutant cAMP receptor protein (CRP*) have been isolated and characterized to be partially relieved from CCR [88-89]. Other *E. coli* strains devoid of the glucose phosphotransferase system (PTS) also resulted in partial elimination of CCR [90]. Deletion of the methylglyoxal synthase gene also modulated the sugar utilization pattern [86a]. Despite several attempts to eliminate CCR, the glucose effect remains a major barrier in cellulosic fuel production.

A yeast strain engineered for cellobiose metabolism was capable of utilizing cellobiose and xylose simultaneously [30]. We observed a similar phenomenon of simultaneous utilization of cellobiose and xylose in the *E. coli* strain ESS that is capable of utilizing cellobiose efficiently. This study investigated the co-utilization of two major sugars of lignocellulosic biomass and its advantages as a host for cellulosic fuel production.

6.3. Materials and methods

6.3.1. Bacterial strains and media

E. coli strains, MG1655, OSS, OSS-pZB-A and ESS were used in this study [91]. Bacterial cells were cultured at 37℃ in Luria-Bertani broth (LB) or M9 minimal medium supplemented with suitable sugars. Strains carrying

temperature-sensitive plasmids were grown at 30℃. For long-term storage, the cells were maintained as 20% glycerol stocks at –80℃.

6.3.2. Batch culture

M9 minimal medium, supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and suitable concentrations of test sugars (cellobiose and xylose), was used to observe co-utilization of cellobiose and xylose. In these tests, overnight cultures grown in LB were inoculated (1:100) into 50 mL of M9 medium supplemented with the test sugars in a 250-mL flask, and the cultures were incubated at 37℃ with rotation at 200 rpm. Cell growth was monitored by measuring the optical density at 600 nm (OD_{600}) with a Biochrom Libra S22 spectrophotometer. Residual sugar was analyzed using high-pressure liquid chromatography (HPLC).

6.3.3. HPLC analysis

The residual sugar in the culture medium during the bacterial growth period was measured regularly using a Shimadzu HPLC station, equipped with an HPX-87P column (Bio-Rad) and a refractive index detector (Shimadzu). HPLC-grade pure water, at a flow rate of 0.6 mL/min, was used as the mobile phase. The oven temperature was maintained at 80℃. A standard curve was determined with varying concentrations of arabinose, galactose, glucose, mannose, xylose and cellobiose.

6.4. Results and Discussion

6.4.1. Glucose-mediated repression of xylose metabolism

In *E. coli*, glucose potentially inhibits xylose utilization. Even though the composition of lignocellulosic biomass is highly variable, the approximate proportion of glucose and xylose in most cases would be 70% and 30%, respectively [30]. Xylose uptake has been reported to be completely inhibited when the molar ratio of xylose: glucose exceeds 1:4 as shown in Figure 6.1 [87]. Therefore, the large excess of glucose in the biomass would inhibit the consumption of xylose during the fermentation process and thus decrease the net productivity.

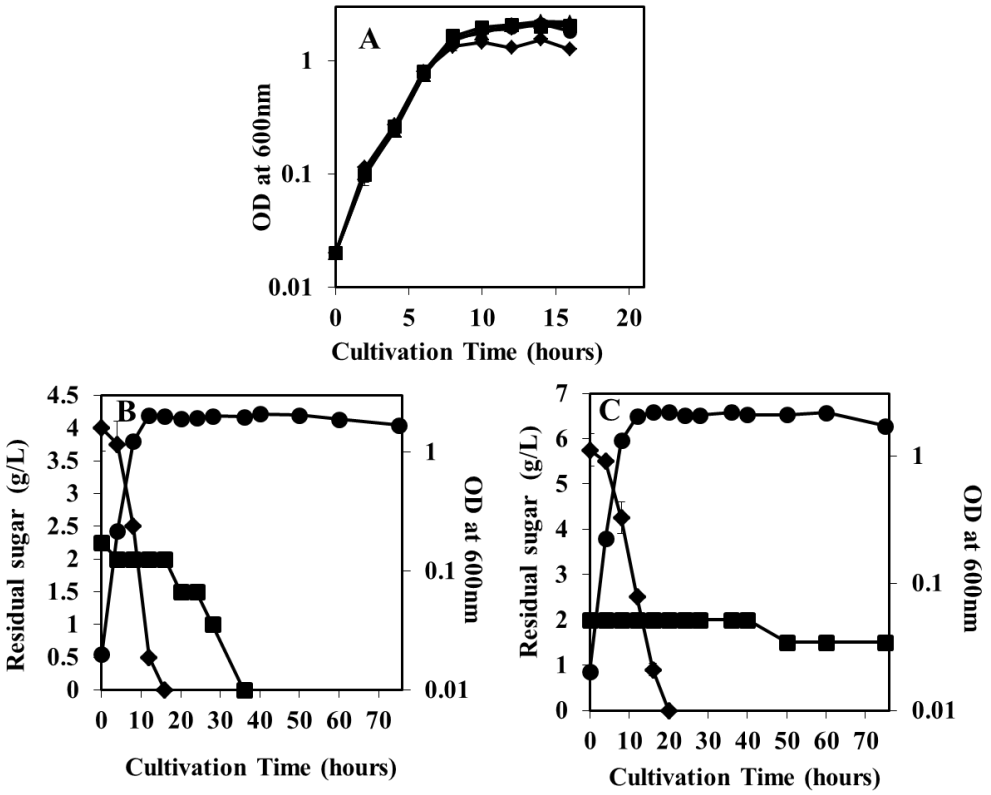


Fig 6.1. Cell growth and substrate utilization pattern of wild type MG1655: (A) Graph showing the effect of higher concentration of glucose on the growth of wild type *E. coli*, MG1655. 2 g/L of glucose (filled diamonds); 4 g/L of glucose (filled squares); 6 g/L of glucose (open plus); 8 g/L of glucose (filled triangles); 10 g/L of glucose (filled circles). (**B** & **C**) Graphs showing the optical density (filled circle), residual concentrations of glucose (filled diamonds) and xylose (filled squares). (**B**) glucose (4g/L) and xylose (2 g/L) and (**C**) glucose (6 g/L) and xylose (2 g/L).

We engineered *E. coli* for the utilization of cellobiose as a sole carbon source through activation of endogenous *chb* and *asc* operons [91]. The resulting strain, ESS, expressed a cellobiose transporter and an intracellular β-glucosidase and was hence capable of transporting cellobiose into the cell and generating glucose within the cell. Cellobiose is a disaccharide of glucose and is hydrolyzed to glucose by the enzyme β-glucosidase. However, the cellobiose-metabolizing *E. coli* ESS was unable to use cellobiose until the complete depletion of xylose. The inhibitory effect observed between cellobiose and xylose metabolism was the reverse of that observed between glucose and xylose metabolism. The xylose-mediated inhibitory effect observed in cellobiose metabolism might not be at the level of transcription or translation, because the cellobiose operon is expressed from a synthetic

constitutive promoter. Hence, the preference for xylose over cellobiose may be attributed to the poor cellobiose-metabolizing capacity of the engineered strain ESS [91].

6.4.2. Simultaneous utilization of xylose and cellobiose by ESS

Previously, an efficient cellobiose-metabolizing strain ESS was isolated after adaptation for 30 days in cellobiose minimal medium [91]. Unlike the unadapted strain OSS, ESS was capable of simultaneous utilization of equal quantities of cellobiose and xylose (Fig. 6.2B). Wild-type *E. coli* was previously reported to sequentially utilize 2 g/L of both glucose and xylose in approximately 20 hours [92]. The ESS strain could simultaneously utilize 2 g/L of both cellobiose and xylose in approximately the same time [30].

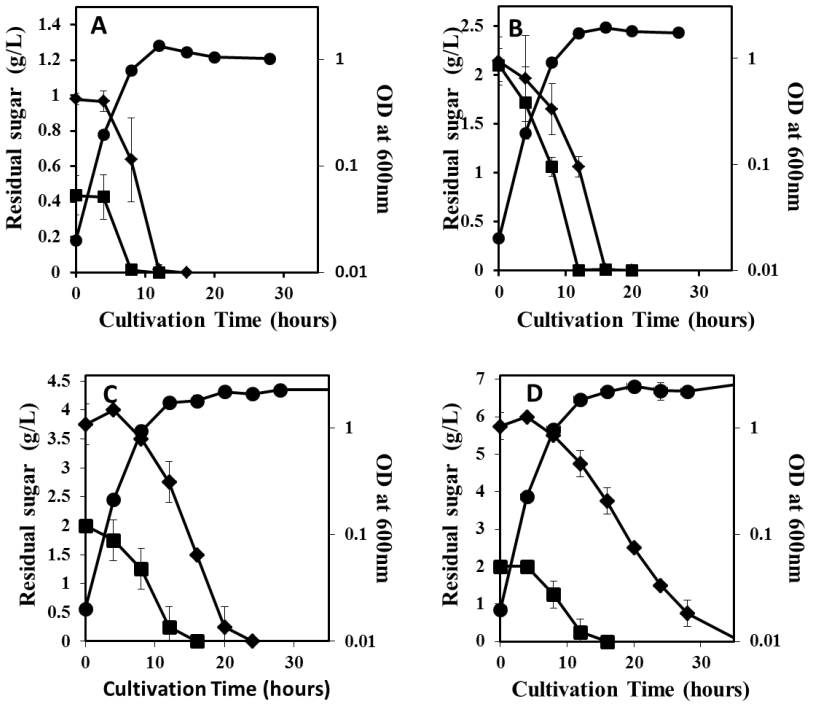


Fig 6.2. Cell growth and substrate utilization pattern of strain ESS: Graphs showing the optical density (filled circles), residual concentrations of cellobiose (filled diamonds) and xylose (filled squares). Strain ESS was grown on minimal medium with varied proportions of cellobiose and xylose: (A) cellobiose (1 g/L) and xylose (0.3 g/L), (B) cellobiose (2 g/L) and xylose (2 g/L), (C) cellobiose (4 g/L) and xylose (2 g/L), and (D) cellobiose (6 g/L) and xylose (2 g/L).

Wild-type MG1655 was unable to completely utilize xylose even after 75 hours when it was grown in a medium containing glucose (6 g/L) and xylose (2 g/L) (Fig. 6.1C) [93]. The ESS strain was capable of complete

utilization of cellobiose (6 g/L) and xylose (2 g/L) in approximately 30 hours (Fig. 6.2C and D), suggesting that co-utilization of cellobiose and xylose might increase the productivity when lignocellulosic substrates are used. A synthetic medium that mimics the palm oil hydrolysate (glucose:xylose ratio of 7:3) [94] was used to determine the efficiency of ESS in a natural hydrolysate (Fig.6. 2A).

In addition to cellobiose/xylose co-metabolism, the strain ESS was capable of simultaneous utilization of galactose or mannose and cellobiose (Fig. 6.3A and B). While xylose is rich in the hemicellulosic portion of hardwood, mannose and galactose comprises the major portion of softwood hydrolysate [87]. Hence, the strain ESS could be a potential candidate for both softwood hydrolysate and hardwood hydrolysate.

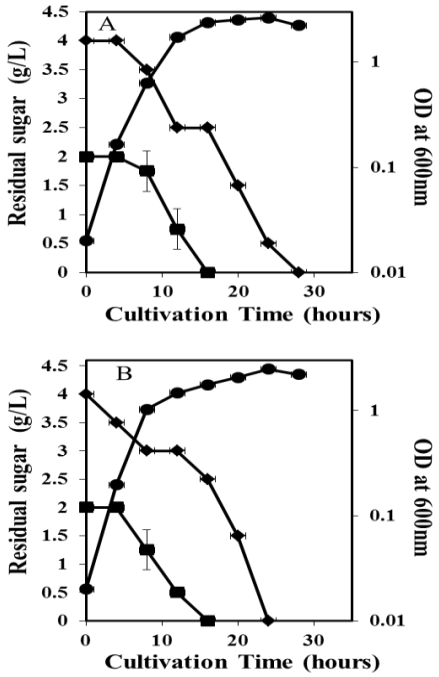


Fig 6.3. Cell growth and substrate utilization pattern of strain ESS: Graphs showing the optical density (filled circles), residual concentrations of cellobiose (filled diamonds) and galactose (A) or mannose (B) (filled squares). Strain ESS was grown on minimal medium with 4 g/L of cellobiose and 2 g/L of galactose (A) or mannose (B).

6.4.3. Role of AscB in co-metabolism

As noted above, the dominant mutation in ESS that affects growth on cellobiose is the mutation in *ascB* gene. We hypothesized that *ascB* overexpression could also favor co-metabolism of cellobiose with other sugars. Due to the increased cellobiose metabolic rate in *ascB* overexpressing cells, the cellobiose consumption rate was

increased. As expected, *ascB* overexpressing strains consumed mannose/galactose and cellobiose more rapidly than OSS (Fig 6.4A2, A3, B2 & B3). These results might indicate that the intracellular glucose produced from the hydrolysis of cellobiose does not contribute to CCR. However, it is likely that the strain overexpressing *ascB* consumed cellobiose after the depletion of xylose unlike ESS (Fig 6.4A1 &B1).

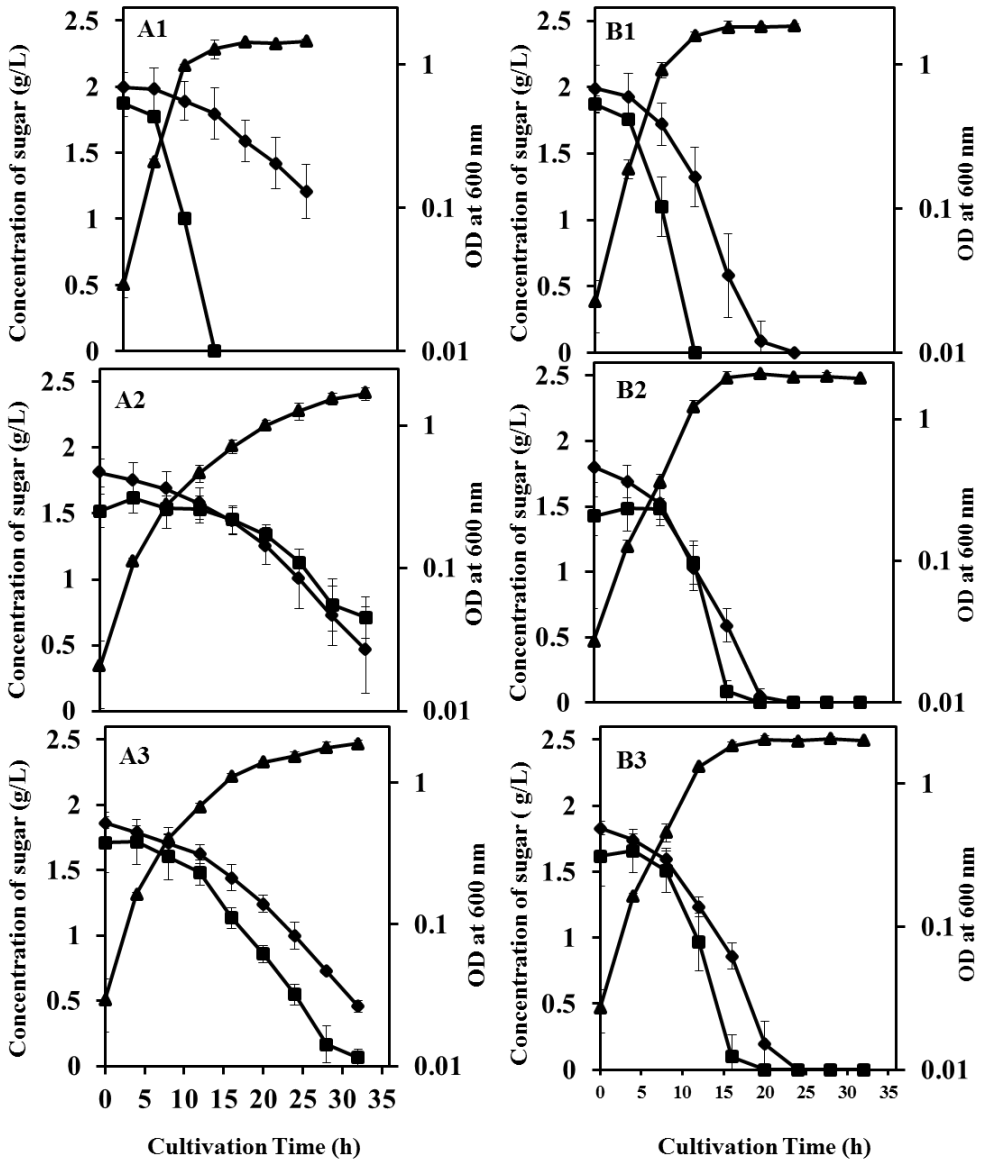


Fig. 6.4. Comparison of co-metabolism of cellobiose (closed diamond) and other sugars (closed square) (1, xylose; 2, galactose; 3, mannose) in OSS (A) and in OSS-pZB-A (B). Closed circles indicate cell growth (OD_{600}). Error bars represent the standard deviation of triplicate experiments.

6.5. Conclusions

To the best of our knowledge, this is the first identified strain of *E. coli* that can efficiently carry out cellobiose and xylose co-metabolism. The adapted strain, ESS, could form a platform strain for simultaneous saccharification and co-fermentation process (SSCF), because it favors simultaneous utilization of the major pentose (xylose) and hexose (glucose in the form of cellobiose) sugars present in hard woody plants. This new strain could also serve as a potential candidate for softwood hydrolysate, because it can utilize cellobiose and mannose simultaneously. Although we used a general lab strain of *E. coli* to show simultaneous utilization of cellobiose and xylose, we believe that this groundwork could be applied to industrial ethanologenic strains or other biofuel-producing strains of *E. coli* to prove its real benefit in the SSCF process.

Chapter 7

Industrial applications of strain ESS

Application II: Imposing cellulolytic ability into *E. coli*

(Engineering *E. coli* as an efficient consolidated bio-processor)

7.1. Abstract

It is necessary to import cellulase system into cellobiose fermenting *E. coli* strains, ESS and OSS, in order to be used for consolidated bioprocessing (CBP). CBP stipulates the need for the expression of a large number of heterologous genes into the microbial host that in turn can impose a huge metabolic burden on the host cell. Therefore, it is necessary to strike a proper balance between the innate metabolic capacity of a microbe and its ability to express heterologous genes. Use of multifunctional cellulases can reduce the metabolic burden during consolidated bioprocessing. Hence, multifunctional cellulases from *Clostridium* (EngB and EngD) and *Bacillus* (Bi-functional cellulase) were expressed in cellobiose fermenting *E. coli* with a view of developing it as an efficient host for consolidated bioprocessing. In addition, transposon mutagenesis was used to identify host factors that are responsible for the repressing the ability of *E. coli* to grow on cellulose. In this study, we present the applications of strain ESS and OSS as a potential consolidated bioprocessor.

7.2. Introduction

It is a long-held belief that multiple cellulases are needed to support the growth of recombinant microbes directly on lignocellulosic biomass (Table 7.1). All the enzymes involved in this process should possess proper synergy and should be well coordinated, such that one step does not subside or supersede other steps. Recent advances in synthetic biology herald the design of artificial microbes and better enzymes for lignocellulose-based biofuel production[95]. Despite several challenges, significant progress has been made with the help of synthetic biology in order to achieve efficient CBP microbes. Reviving cellulase expression and secretion system is the most challenging step in designing industrial cellulolytic organisms. Different types of cellulases should either be secreted extracellular (free cellulases) or bound to the surface (cellulosomes) for the efficient hydrolysis of lignocellulosic biomass. Combinations of over 400 different cellulose-related enzymes are present in well-known cellulolytic organisms including *Trichoderma*, *Aspergillus*, *Cellvibrio*. Most of these enzymes should be introduced into an ideal CBP microbe for the efficient depolymerization of lignocellulosic residues and their subsequent conversion to simple sugars. The expression and secretion of multiple genes is a prerequisite for efficient growth on lignocellulosic biomass. Cellulases with higher activity, wide substrate specificity, thermo-tolerance, and chemical tolerance, obtained either from metagenomic libraries or from rational protein evolution, are aimed at reducing the minimal set of cellulases needed for ideal CBP strains[96 96b].

Cellulase secretion is a long-standing constraint in developing an ideal CBP microbe. Cellulose, being a larger molecule, demands an ideal CBP organism to efficiently secrete different cellulases into the medium. Practically, it is necessary to secrete as high as 20 mg cellulase/g of solid cellulosic material[97]. Several attempts have been made to either secrete free cellulases or display cellulosomes on the surface of recombinant microbes. For example, the use of fusion proteins like OsmY or increasing membrane porosity *via* the deletion of lipoproteins increased the extracellular titer of heterologous proteins to ~70 mg/L in *E. coli*[98]. Despite several attempts, sufficient protein secretion for efficient growth on cellulose could not be achieved probably because of the high energy consumption associated with this process. For example, the translocation of a single protein through the SecB or TAT pathway requires 10^5 protons (~10^4 ATP molecules)[99]. Surface-display technology has helped in mimicking the cellulosomal structures observed in anaerobic cellulolytic bacteria[100]. The cellulosome offers 2 main advantages over free cellulases: they do not diffuse out into the surrounding media and the close proximity of different cellulases help enhance their synergy to hydrolyze biomass. In this study, we expressed different

cellulolytic enzymes into strain OSS and ESS in order to make it proficient for consolidated bioprocessing.

Table 7.1. List of enzymes needed to make the complete cellulolytic cocktail for the saccharification of lignocellulosic biomass

Enzymes	Cleavage target	Abundance
Cellulases		
Cellobiohydrolase	Hydrolyzes reducing or non-reducing end of cellulose chain	All plant biomass
Endoglucanase	Hydrolyzes random internal glycosidic bonds of cellulose	All plant biomass
β-glucosidase	Cleaves non-reducing end of cellobiose and cellodextrin	All plant biomass
Phospho-β-glucosidase	Cleaves phosphorylated cellobiose and cello-dextrins	All plant biomass
Hemicellulases		
Endoxylanase	Hydrolyzes glycosidic bond of glucuronoxylan or arabinoxylan	Hardwood, grasses and cereals
β-xylosidase	Hydrolyze xylobiose or xylooligosaccharide from xylan	Hardwood, grasses and cereals
Acetyl xylan esterase	Deacetylates the backbone glycosyl unit of xylan	Hardwood xylan
Feruloyl esterase	Hydrolyzes feruloyl esters of arabinan/arabinoxylan, rhamnogalacturonan or xyloglucan	Grasses and cereals
Glucuronoyl esterase	Demethylates xylan backbone of glucuronoarabinoxylan	Grasses, cereals and softwood
Arabinofuranosidase	Removes arabinose from arabinoglucuronoxylan or arabinoxylan	Softwood and grasses
Galactosidase	Removes galactose from galactomannan or pectin	Softwood
Glucuronidase	Removes glucuronoyl or its methyl ester from xylan	Hardwood
Mannanase	Degrades mannosyl polymer of glucomannas or galactomannas	Softwood and hardwood
Xyloglucan Hydrolase	Degrades glucan of xyloglucan	Hardwood and grasses
Pectinolytic enzymes		
Polygalacturonases	Cleaves glycosidic bond of polygalactouronan	Sugar beet pulp and fruit
Pectin/Pectate Lyases	Cleaves O-C_4 glycosidic bond of polygalactouronan	Sugar beet pulp and fruit
Pectin Methyl Esterase	De-esterifies pectin	Sugar beet pulp and fruit
Lignin Degradation		
Lignin peroxidase	Oxidatively degrades lignin	All plant biomass
Aryl-alcohol oxidase	Oxidation of aromatic alcohol of lignin	All plant biomass
Laccase	Degrades phenolic portion of lignin	All plant biomass
Glyoxal oxidase	Oxidation of glyoxal	All plant biomass
Cellobiose dehydrogenase	Oxidize cellobiose and cellodextrins to aldonolactones	All plant biomass
Cell wall loosening enzymes		
Expansin	Expansion, slippage or lengthening of cell wall structure	All plant biomass
Swollenin	Promotes breakdown of cellulose	All plant biomass
Loosinin	Increases cellulase accessibility	All plant biomass
Cellulose induced protein	Cleaves hemicellulose-lignin cross-link	All plant biomass

7.3. Materials and methods

7.3.1. Bacterial strains and media

Strains and Plasmids used in this study are listed in Table 7.2. Bacterial cells were cultured at 37℃ in Luria-Bertani broth (LB) or M9 minimal medium supplemented with 1g/L of cellulose and 1g/L yeast extract. For long-term storage, the cells were maintained as 20% glycerol stocks at –80℃. M9 minimal medium, supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.05 g/L yeast extract and suitable cellulose, was used to observe growth on

cellulose. In these tests, overnight cultures grown in LB were inoculated (1:100) into 50 mL of M9 medium in a 250-mL flask, and the cultures were incubated at 37℃ with rotation at 200 rpm. Cell growth was monitored by measuring the optical density at 600 nm (OD_{600}) with a Biochrom Libra S22 spectrophotometer or by plating desired dilutions in LB plates.

Table 7.2. Strains and plasmids used in this study.

Strains/plasmids	Description/genotype	Reference/source
Strains		
E. coli MG1655	Wild type	[37]
MG1655/Δ*lpp*::*frt*	MG1655 with *lpp* gene deleted	This study
OSS	MG1655 with CP12chb plus CP12asc	[18]
ESS	OSS adapted in cellobiose for 30 days	[18]
Plasmids		
pL-B	pL31b+ vector expressing Bi-functional cellulase from Bacillus	This study
pET-B	pL31b+ vector expressing Bi-functional cellulase from Bacillus	This study

7.3.2. Construction of Transposon Library

Random gene knockouts in strain OSS was obtained by electroporation of the EZ::TN<R6K ori/KAN-2 along with the Type I restriction inhibitor, recovered for 1 hour in SOC medium and plated on LB-Kanamycin plate. 6000 colonies carrying random transposon insertion were pooled into 4 mL LB (1500 CFU/mL) and stored as 20% glycerol stock at -80℃. 200 µL of transposon library (stored as glycerol stock) (300 CFU) was transformed with plasmid pL-B and recovered in SOC medium for 1 hour and 30 minutes. Cells were then transferred to 50 mL LB broth containing Amp and Kan. Cells were grown for 5 hours till the OD becomes ~0.25. Cells were then collected by centrifugation and suspended in M9 minimal medium with 0.1% CMC and 0.005% yeast extract to a final OD of 1 and enriched 3 times on CMC supplemented minimal medium. The transposon insertion sites in the final clonal population were verified by rescue cloning of *EcoRI* digested genomic DNA or through RATE PCR.

7.3.3. Enzyme assay

1% of overnight grown cells were diluted into LB containing the suitable antibiotics and grown at 37℃ at 200

rpm. Cells were induced with 100 µM IPTG and grown at 30℃ for 4 hours. 2 mL cells were collected and fractionated into periplamic and cytoplasmic fraction. Briefly, for periplasmic fraction, cells were suspended in tris (300 mM)-sucrose (20%)-EDTA buffer and vortexed slowly and centrifuged. The supernatant was discarded and the pellet resuspended in ice-cold MgSO4 (5 mM) and centrifuged to extract the periplasmic fraction. To the resulting pellet, tris-lysozyme was added and sonicated to extract the soluble fraction. For enzyme assay, 100 µL of the cellular extract was mixed with 1% CMC in sodium phosphate buffer and incubated at 37℃ for 2 hours. The hydrolyzed sugar concentration was analyzed using DNS-reducing sugar assay.

7.4. Results and Discussion

7.4.1. Verification of functional expression of multi-functional cellulases in *E. coli*

Bi-functional cellulase from *Bacillus* is a multi-functional cellulase (with both endo and exoglucanase activity) capable of producing cellobiose directly from cellulose without the need for additional exoglucanase[97]. Functional expression of the multifunctional cellulase was verified by expression from pET promoters in BL21 (DE3) (Fig 7.1A). Despite higher portion being retained in the insoluble fraction, bi-functional cellulase retained the native cellulolytic activity when tested against carboxymethyl cellulose (CMC) (Fig 7.1A). In order to favor expression in cellobiose metabolizing *E. coli*, pET promoter was replaced with P_L promoter and tested for its activity (Fig 7.1B). Bi-functional cellulase was expressed in active/soluble form both from P_L and pET promoters. Hence, expression of bi-functional cellulase in OSS or ESS should allow these strains to grow on cellulose directly without the addition of β- glucosidase. Bi-functional cellulase was used to analyze its efficiency in consolidated bioprocessing in strain OSS or ESS. Even after 72 hours of cultivation there was no significant growth observed on soluble cellulose or a change in cell number on insoluble cellulose. Even when the initial cell OD was increased or yeast extract supplemented to CMC-Na, there was no significant increase in growth both in strain OSS and in strain ESS. Hence, the host factors contributing to the inability of strain, OSS for growth on cellulose was analyzed using transposon mutagenesis.

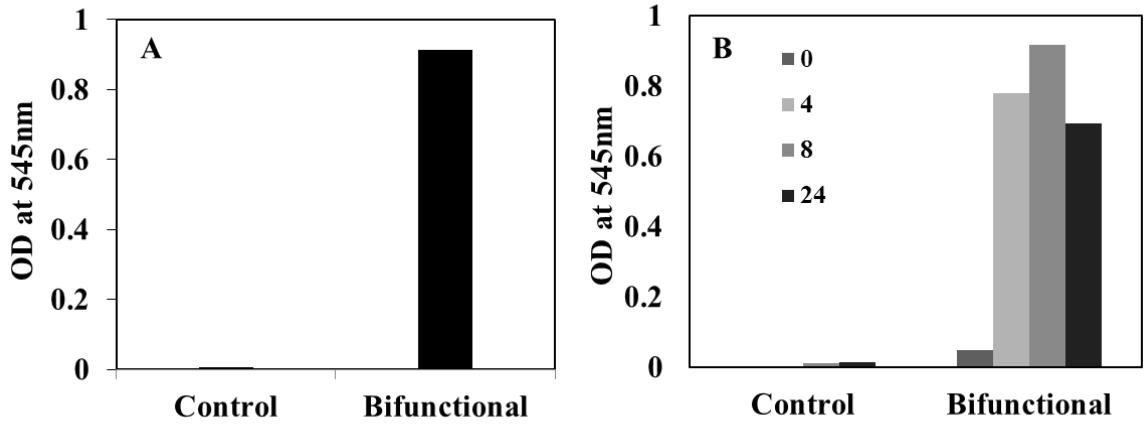


Fig 7.1. Enzyme activity of Bifunctional cellulase against CMC upon expression in *E. coli* BL21(DE3) from pET plasmid (A) or pL plasmid (B).

7.4.2. Enriching transposon library on soluble cellulose

Since both strains ESS and OSS (with different growth rates on cellobiose) could not allow growth on cellulose upon expressing bi-functional cellulase, there could be other host factors that is controlling the co-ordination between cellulase secretion, cellulose hydrolysis and cellobiose metabolism. Hence, random gene knockout library was constructed in strain OSS. Approximately 6000 random gene knockout library was constructed in strain OSS and then transformed with pL-B plasmid. The pL-B plasmid carrying OSS mutant library was then enriched on cellulose and screened for fast growing strains on cellulose. After three successive enrichments on cellulose, strains were plated on LB-plates. Ten independent clones were selected and analyzed for growth on cellulose. Three representative clones had a significant increase in growth on cellulose (Fig 7.2.). The genes inactivated in the identified clones are listed in Table 7.2. Interestingly, the clones obtained from transposon library had a reduced enzyme activity compared to the wild type strains expressing the same enzyme (Fig 7.3). In addition, growth of strains OSS or the transposon clones (enriched on cellulose) on cellobiose was not altered indicating that cellobiose metabolism may not be the rate-limiting step for growth on cellulose.

Genes related to ribosome or extracellular polysaccharide was inactivated with the transposon to allow growth on cellulose. Analysis of the impact of these genes on cellulase expression indicate that the deletion of these genes (particularly *rfbA* and *tufA*) leads to a reduced cellulase activity. It could be possible that the mutations in *rfbA*

modifies the surface polysaccharide and enhancing the protein secretion. The *tufA* gene is expressed at a higher level during stress response and is one of the most abundant protein in *E. coli*. Thus, *tufA* and *rfbA* gene products encode for essential cellular process. Noteworthy is both *rfbA* and *tufA* genes have a functional paralogs in *E. coli* indicating that the mutations obtained through transposon mutagenesis is not deleterious for host process. These results might also indicate that a balanced cellulase expression might be one of the dominant factors to be optimized to allow strain ESS and OSS on cellulose even without the addition of extra β-glucosidase. In line with that all clones exhibiting improved growth on cellulose exhibited a reduced enzyme activity compared to wild type strains (Fig 7.3).

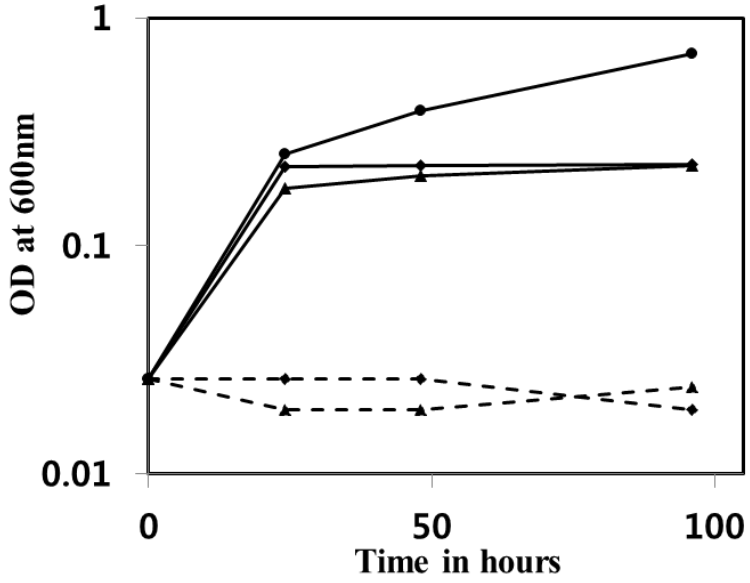


Fig 7.2. Growth of strains expressing bi-functional cellulase on soluble cellulose with yeast extract. ESS with pL-B, *solid triangle*; OSS with pL-B, *solid diamond*; OSS-TN5-clone 9 with pL-B, *solid circle*. Dashed lines indicate OSS (solid diamond) or ESS (solid triangle) carrying empty pL plasmid.

Table 7.2. Genes inactivated in transposon library enriched on soluble cellulose.

Strain number	Gene inactivated	Gene Function
Clone 6	*rrlG*	23S rRNA, component of ribosome
Clone 8	*rfbA*	Synthesis of essential component of surface polysaccharide
Clone 9	*tufA*	Ribosomal protein, expressed during stress response

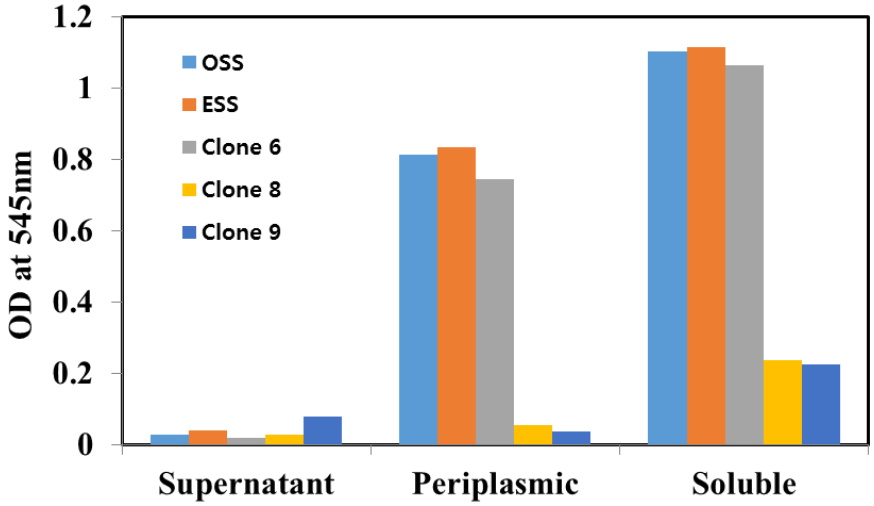


Fig 7.3. Comparison of pL-B activity in strains OSS, ESS, and clones 6, 8, and 9 obtained from transposon library.

7.5. Conclusions

Optimizing the cellulase expression in strains, ESS or OSS would help in obtaining an efficient consolidated bioprocessor that mimics the native cellulolytic microbes. This process would also help in reducing the need for secretion of extra β-glucosidase as protein secretion is a high energy demanding process. This study also clearly indicates the need to optimize the cellulase expression level rather than to increase the expression level of the enzyme as a whole. An optimized strain expressing the cellulase and the cellobiose metabolic pathway at an efficient-level would be a next step towards consolidated bio-processing. The strain ESS or OSS could thus serve as a starting point for developing a strain for consolidated bioprocessing.

Appendix I: List of genes that are significantly up/down-regulated in ESS when compared to OSS grown on cellobiose minimal medium. Strains were grown on M9-cellobiose minimal medium to an optical density of 0.7 at 37℃ with 200 rpm. The values in column for test 1, test 2, and test 3 represent the signal intensity value obtained from microarray data. The column test corresponds to the average of test 1, test 2 and test 3. The column ratio corresponds to the fold change in gene expression between strain ESS and OSS. The summary of changes are described in chapter 3.

Name	test				test_1		test_2		test_3	
	Ratio	t-test P-value	ESS	OSS	ESS	OSS	ESS	OSS	ESS	OSS
fliC	50.78	1.45E-02	907	25	1061	56	864	10	798	10
pyrI	25.98	1.13E-06	8432	322	8676	354	10005	338	6616	273
pyrB	23.91	4.32E-04	7101	299	7797	343	6640	306	6865	248
flgE	18.62	1.19E-02	936	60	1197	106	813	51	799	22
fiu	10.97	2.08E-06	2769	252	3532	316	2530	233	2246	206
flgB	10.88	2.11E-05	653	62	796	95	732	53	431	39
fliZ	10.47	4.52E-03	504	51	553	80	453	35	507	39
flgD	9.41	3.64E-03	1522	170	1725	251	1519	139	1321	120
malE	8.26	1.74E-03	1013	123	1098	168	1232	106	709	96
fliS	7.78	2.45E-02	179	27	159	48	258	22	120	10
malP	7.56	1.47E-06	552	77	928	136	385	58	342	36
codA	7.36	1.29E-05	781	107	927	127	912	128	504	66
fliK	7.32	2.38E-03	238	35	319	59	197	27	198	20
fliA	7.24	4.24E-03	823	124	1157	210	688	96	624	65
fepA	6.62	3.72E-04	8819	1324	8384	1277	10621	1518	7451	1178
upp	6.51	3.20E-03	1172	185	1198	237	1240	187	1078	131
gdhA	6.34	4.27E-07	2420	378	2676	411	2696	394	1887	330
gadA	6.05	5.78E-03	485	81	468	91	434	91	553	62
yjgF	5.99	4.17E-09	1417	239	1623	299	1519	241	1108	176
rutG	5.78	6.77E-02	84	10	131	10	14	10	108	10
hsdS	5.76	7.16E-03	164	29	142	37	196	29	156	21
pyrC	5.74	3.39E-04	2480	432	3096	523	2252	429	2093	344
ilvC	5.34	5.24E-05	4023	748	4666	811	4206	791	3198	642
hdeB	5.21	1.58E-08	3039	589	3784	770	2906	558	2427	439
murI	4.94	2.86E-02	97	23	109	43	52	10	131	14
guaA	4.59	8.30E-13	3838	833	4076	892	4253	889	3185	720
asd	4.59	5.07E-04	1518	328	1876	370	1431	333	1248	281
ascB	4.56	8.97E-12	41842	9000	61090	12117	36113	8351	28322	6533
cyoB	4.50	1.17E-09	2323	517	2424	560	2488	554	2057	437
waaA	4.50	2.97E-06	109	25	126	28	119	32	81	15
dinI	4.41	2.28E-03	465	107	441	133	536	105	416	83

hdeA	4.40	1.70E-04	5101	1151	5280	1210	6187	1336	3837	909
yfeK	4.32	2.15E-02	78	14	161	23	50	10	24	10
lysC	4.24	1.31E-05	1215	279	1529	333	1374	297	742	207
gspE	4.11	no replicates	798	194	798	194				
ychM	4.06	1.54E-01	167	22	63	32	19	10	419	22
pdhR	4.01	8.76E-11	824	210	1027	288	755	195	690	148
dcp	3.95	2.07E-03	313	82	303	103	350	89	286	54
metE	3.94	1.38E-04	3900	991	4249	1093	3923	1009	3529	871
csgF	3.94	5.15E-02	81	23	28	10	85	10	129	50
oppB	3.91	4.94E-03	159	46	165	60	212	61	100	16
gadB	3.84	2.13E-02	538	137	526	147	706	115	382	148
ompT	3.82	1.37E-03	19514	5019	25890	5969	18230	5026	14422	4061
fimI	3.76	3.40E-04	475	125	400	119	554	137	471	121
cyoA	3.73	1.44E-08	3266	875	3320	942	3706	940	2773	743
atpB	3.72	3.25E-04	1610	436	1730	494	1724	473	1376	342
fecE	3.71	1.00E-02	184	50	134	62	263	58	155	30
flhD	3.70	5.47E-06	362	103	638	194	229	65	219	50
pntB	3.66	1.94E-04	2150	587	2018	667	2500	587	1931	507
nrdB	3.63	7.67E-03	329	104	584	217	227	55	176	41
guaB	3.62	8.92E-04	1559	418	1359	429	2155	457	1164	367
ybbL	3.58	1.15E-02	165	44	133	60	269	44	94	28
yfdH	3.57	4.56E-04	708	200	839	242	677	202	609	155
marR	3.56	2.60E-02	90	23	35	16	93	34	143	18
flgL	3.55	4.57E-07	268	77	279	81	267	90	259	59
aroG	3.53	9.01E-04	4152	1186	4030	1252	3954	1329	4471	977
entD	3.47	4.89E-08	1471	420	1513	464	1751	435	1149	359
mrdA	3.38	1.19E-03	194	60	227	85	216	63	141	34
yeiB	3.35	3.31E-03	366	112	302	121	277	63	519	151
pyrG	3.35	4.21E-10	1117	338	1246	423	1096	324	1009	268
cusF	3.33	3.97E-02	187	42	63	27	31	10	469	89
insL	3.32	2.64E-02	128	38	86	39	158	30	139	44
ybdB	3.30	5.37E-04	3107	951	3214	1159	3408	920	2701	773
fimA	3.30	9.88E-04	1889	572	1894	580	2154	629	1619	506
ybdM	3.29	2.83E-02	203	119	472	329	90	17	48	10
gltD	3.28	9.68E-04	2482	759	2987	921	2265	715	2195	641
aceE	3.27	4.29E-05	2112	638	2319	739	2483	640	1533	533
purD	3.26	2.02E-03	1194	369	1621	519	944	291	1017	298
dapD	3.25	3.83E-04	989	308	915	359	1204	346	847	219
insD	3.25	6.22E-07	577	177	499	175	674	189	560	166
atpE	3.21	4.69E-04	4957	1523	5895	1728	5454	1603	3523	1237
atpH	3.20	4.81E-05	3412	1076	3656	1279	3400	1059	3182	889
ppc	3.19	5.84E-07	2137	667	2158	755	2651	746	1602	500
livJ	3.19	1.38E-03	8114	2547	8936	2825	8423	2674	6983	2143

atpF	3.19	8.68E-09	3874	1205	4373	1344	4286	1248	2964	1023
proX	3.19	8.49E-03	340	110	335	150	368	90	318	90
accB	3.19	3.17E-03	3121	974	4436	1336	2805	928	2122	658
serA	3.17	7.93E-08	5219	1614	6860	1827	4900	1682	3899	1333
glgB	3.14	8.06E-04	1763	556	2314	711	1773	542	1202	414
lamB	3.14	1.24E-04	891	283	899	328	1059	281	716	240
ptsH	3.12	2.85E-08	3428	1110	3888	1372	3423	1086	2972	870
hisG	3.12	3.62E-07	1005	322	962	349	1099	327	955	291
ubiX	3.06	1.01E-01	59	27	70	60	33	10	75	10
insD	3.03	4.68E-03	514	161	493	139	752	204	297	140
imp	3.02	5.08E-04	705	226	668	225	946	253	501	201
moaD	3.01	1.08E-03	694	229	653	248	607	203	821	236
yidD	3.01	8.20E-06	551	183	600	204	546	169	506	176
cyoE	3.00	3.92E-04	2183	731	2354	912	2499	701	1697	579
flgG	2.98	1.44E-06	476	158	379	139	589	184	461	152
hisC	2.98	2.27E-03	580	197	754	266	545	188	441	137
mprA	2.97	5.27E-04	542	182	993	335	369	98	264	112
glmU	2.97	1.20E-03	656	225	691	271	648	240	629	165
yfdI	2.96	9.43E-04	958	325	967	369	1056	339	852	268
mutT	2.96	2.20E-02	108	38	125	30	102	30	98	52
cusB	2.96	2.28E-02	212	79	542	207	31	19	62	10
acpT	2.94	4.28E-02	160	60	158	92	144	30	178	57
pqqL	2.94	2.33E-08	264	90	272	104	269	90	250	77
adk	2.93	3.00E-04	2493	826	3405	971	2404	839	1671	669
insD	2.92	1.36E-07	510	174	396	138	649	220	484	165
kdgT	2.91	6.86E-02	167	52	260	44	151	57	90	57
insD	2.90	7.25E-05	546	186	440	180	620	193	578	185
pyrF	2.90	2.78E-02	275	123	432	267	215	56	179	46
purH	2.90	2.68E-03	915	315	911	334	1020	333	815	279
borD	2.89	1.62E-02	164	57	167	69	164	52	162	51
yadG	2.88	4.19E-02	201	73	190	93	186	82	226	44
hsdM	2.88	2.51E-03	279	95	352	88	292	118	195	80
pgi	2.88	1.04E-07	919	326	1079	438	896	287	782	253
rbsR	2.86	2.24E-07	1740	623	2515	949	1527	541	1178	379
potC	2.85	9.81E-05	141	51	132	57	137	34	155	63
fliL	2.80	4.78E-02	345	136	338	198	402	149	296	62
atpA	2.78	1.49E-03	3962	1422	4291	1635	4378	1428	3217	1204
rpoB	2.76	1.65E-03	1670	603	1709	646	1936	660	1364	501
folE	2.76	6.44E-10	2004	726	2309	804	1980	772	1722	601
betB	2.75	5.29E-07	2074	757	2299	923	2310	774	1613	574
ilvI	2.75	1.73E-03	941	344	1075	405	934	356	813	272
yfjH	2.75	1.61E-01	954	167	2725	416	58	63	79	23
fbaA	2.74	1.77E-06	3279	1202	4074	1529	3067	1102	2697	974

tig	2.73	1.82E-03	2770	1010	2752	1108	3343	1085	2215	836
gadE	2.72	4.80E-05	362	133	328	135	338	113	420	153
yejK	2.70	1.39E-04	321	121	345	154	303	118	317	92
gcvT	2.70	5.65E-03	296	113	279	79	235	96	374	164
katG	2.70	1.07E-01	287	107	188	130	306	127	367	65
flgC	2.69	1.21E-03	1211	454	1492	607	1211	349	930	406
entC	2.69	4.43E-06	1844	697	2136	853	1782	731	1615	506
ghrB	2.68	2.14E-03	864	339	1359	589	612	226	622	202
flgH	2.68	1.88E-03	295	101	506	136	213	90	167	76
efeO	2.67	2.10E-03	1564	586	1391	597	1775	648	1525	513
recB	2.66	2.55E-02	297	115	320	158	283	86	288	101
yffM	2.66	9.72E-02	53	19	17	13	66	10	78	36
ypjA	2.66	7.14E-04	116	50	194	101	87	24	66	24
ycbT	2.66	7.78E-05	152	58	173	77	146	48	137	50
yncE	2.65	6.44E-04	4749	1744	6467	1885	4751	2068	3029	1279
pdxH	2.63	8.44E-04	498	191	483	236	565	174	446	162
yfdR	2.62	no replicates	149	57					149	57
bcsF	2.61	1.31E-02	301	77	761	164	45	30	97	38
speD	2.61	1.46E-05	568	218	702	276	452	196	551	182
stpA	2.61	1.82E-07	242	94	286	119	250	97	191	67
fliF	2.61	1.04E-03	253	101	224	70	254	92	280	140
rph	2.60	1.63E-02	254	95	281	128	327	84	153	74
maeA	2.59	4.03E-06	882	342	907	385	993	372	745	270
insD	2.59	2.47E-03	2160	846	2681	1119	1888	770	1912	650
sufA	2.58	2.50E-03	806	313	837	337	838	326	744	276
gapA	2.58	2.22E-03	5893	2244	7734	2577	5558	2347	4389	1809
proV	2.58	2.44E-08	265	103	309	123	253	93	233	93
tbpA	2.58	2.94E-03	452	173	536	175	450	172	371	173
tpiA	2.57	1.94E-07	2374	919	2390	1053	2886	973	1848	733
yibQ	2.56	8.87E-04	195	78	333	138	122	49	129	46
yhbJ	2.56	8.47E-07	835	329	892	381	810	336	802	271
yqfE	2.55	9.38E-02	76	46	29	12	58	10	141	116
coaA	2.54	2.26E-02	219	85	151	93	276	101	231	62
betI	2.54	2.89E-03	2545	1031	3122	1464	2695	960	1817	668
yciI	2.53	2.23E-06	549	218	598	251	548	215	500	187
rplK	2.52	2.84E-03	7088	2729	9732	3422	7073	2699	4459	2066
cyoC	2.52	2.03E-03	1145	455	1225	494	1153	420	1058	451
yfdG	2.51	4.07E-05	305	123	369	135	327	151	218	81
purU	2.51	3.23E-03	481	193	506	223	441	178	496	177
holC	2.50	5.48E-03	494	201	723	310	368	154	390	139
ykfJ	2.50	1.13E-01	717	304	1070	697	1060	204	20	10
yadH	2.49	8.52E-02	223	97	191	138	250	102	228	50
intF	2.49	1.00E-02	182	73	172	85	172	72	203	63

fhuE	2.49	1.36E-03	2488	961	4107	1466	1424	758	1934	659
pyrH	2.49	2.19E-03	837	331	1104	377	633	345	775	270
scpB	2.47	1.35E-01	37	16	32	22			41	10
insD	2.46	3.31E-03	439	177	369	169	527	195	422	167
yfdP	2.45	2.03E-01	143	119	277	329	126	18	25	10
cirA	2.45	1.08E-06	13339	5426	13863	6153	15566	5686	10589	4440
glnS	2.45	3.25E-06	1080	442	1311	546	1094	438	836	341
ybiC	2.45	6.73E-03	574	237	539	279	666	251	518	181
yhjE	2.44	4.06E-03	640	265	677	300	666	277	578	216
yojI	2.42	6.49E-03	245	100	199	101	279	98	256	101
yfbU	2.42	1.47E-01	1179	288	2955	473	303	222	280	169
yegK	2.42	5.21E-03	1237	509	1175	529	1038	564	1500	434
mreD	2.42	1.61E-03	430	173	551	195	294	158	444	165
malM	2.41	2.21E-02	487	202	558	150	553	253	350	202
ptsI	2.41	5.70E-07	2451	1022	2738	1248	2657	1011	1958	806
gsk	2.40	2.89E-03	195	83	233	99	178	96	174	55
plsC	2.39	2.23E-02	177	72	203	92	104	60	226	63
yddA	2.38	9.44E-02	72	36	67	24	30	10	119	73
pnp	2.38	1.58E-03	2179	916	2186	1023	2441	941	1910	783
rluF	2.35	5.83E-03	297	127	420	184	219	104	254	93
ampC	2.34	2.27E-02	224	137	547	375	22	10	103	26
minC	2.33	1.02E-02	1087	466	1451	636	1030	411	780	352
yafJ	2.32	1.69E-03	374	164	589	267	288	133	244	93
fliG	2.32	2.39E-03	474	217	680	368	320	145	422	139
nrdF	2.32	5.82E-03	538	232	705	337	604	214	305	145
livK	2.31	1.03E-06	2533	1094	2821	1248	2731	1115	2046	920
ycjS	2.30	1.51E-02	1390	606	1094	662	1774	729	1302	427
ydiT	2.30	2.91E-01	108	37	42	34			174	41
ydcX	2.29	3.53E-02	171	68	81	57	177	75	256	70
bcp	2.28	2.37E-03	1115	482	877	506	1491	564	978	376
atpG	2.28	5.23E-03	2473	1079	3107	1322	2389	1016	1922	898
fhuF	2.28	9.08E-07	3230	1410	3538	1653	3739	1438	2414	1138
valS	2.27	4.72E-03	868	379	884	398	1038	418	683	321
lptA	2.27	3.87E-03	1399	591	1940	708	1410	597	847	467
sthA	2.26	5.68E-03	1327	578	1690	722	1432	587	860	425
insD	2.25	4.50E-03	359	156	263	162	401	145	412	162
ushA	2.25	1.41E-05	300	136	379	177	264	141	257	91
insC	2.25	1.34E-02	700	313	732	351	787	342	582	245
bioD	2.23	1.16E-02	178	80	170	64	223	105	140	70
rhlB	2.23	7.19E-03	1409	633	1405	694	1376	650	1445	556
gltA	2.22	6.32E-03	9510	4305	9702	4947	9428	4194	9400	3775
aceF	2.22	6.25E-03	2911	1306	3483	1408	2619	1388	2632	1124
rnpA	2.21	3.29E-03	685	307	923	390	510	306	621	225

speE	2.20	7.04E-03	1266	583	1270	703	1483	628	1045	417
uhpC	2.20	1.66E-01	777	314	2080	799	105	109	145	34
bcsE	2.20	8.92E-03	223	100	156	78	291	124	221	97
xylR	2.20	5.76E-02	158	71	325	147	88	33	60	33
eno	2.19	8.48E-06	3682	1669	4039	1852	4112	1734	2895	1421
sucD	2.19	1.60E-05	4975	2253	6125	2744	5240	2262	3558	1753
ynfC	2.18	3.51E-02	175	80	151	91	217	60	156	90
ompC	2.18	8.10E-03	5702	2621	5615	2853	5870	2689	5622	2320
hmp	2.17	9.94E-02	125	66	110	70	111	25	154	104
rpmH	2.17	7.44E-03	1121	505	1400	550	1196	558	766	408
flgN	2.17	9.26E-03	734	367	1680	871	268	131	254	98
hpf	2.17	7.30E-03	2114	978	2210	1092	2295	1019	1838	825
rpsR	2.16	9.51E-06	7222	3326	8098	3599	7641	3524	5926	2856
fhuA	2.16	6.87E-03	1281	596	1512	683	1281	644	1052	460
ygcW	2.16	2.46E-01	71	39	97	84	25	19	90	14
dnaG	2.16	1.11E-02	1051	495	974	564	1019	549	1161	372
yagU	2.15	4.48E-06	721	337	779	394	668	322	716	296
pepP	2.15	1.02E-06	606	282	732	335	552	268	533	242
yheS	2.15	1.05E-02	268	127	351	183	203	94	249	104
ynfO	2.15	7.81E-02	92	45	50	35	144	77	83	22
btuB	2.15	1.07E-02	584	275	641	326	598	279	513	218
pepA	2.13	3.93E-02	183	97	308	180	123	79	120	33
mglA	2.13	8.32E-03	599	282	543	281	696	326	560	238
ghrA	2.13	8.36E-03	299	140	259	128	334	143	304	149
hupA	2.11	1.52E-02	2695	1251	3834	1630	2312	1144	1940	980
dacA	2.11	1.05E-02	318	150	332	169	253	128	370	154
serC	2.11	3.03E-05	3154	1476	3771	1658	3460	1584	2231	1186
rplA	2.11	8.91E-03	5591	2692	6193	3252	5623	2770	4957	2054
prfA	2.10	1.05E-02	309	146	361	166	226	151	341	120
pgk	2.10	3.77E-03	4199	2010	4736	2218	4053	2140	3809	1672
flgM	2.10	1.15E-02	306	141	216	121	426	164	275	139
fecA	2.09	2.07E-05	499	238	563	264	496	243	438	207
yfdX	2.09	2.92E-02	181	60	453	122	51	29	39	28
gyrB	2.09	1.09E-02	317	147	307	159	225	142	421	142
tktA	2.08	1.45E-05	1864	900	1850	1039	2119	930	1623	732
entA	2.08	2.33E-03	3797	1829	3763	2186	4601	1859	3027	1441
yehL	2.07	7.39E-03	524	260	529	342	525	249	517	189
glgA	2.07	1.07E-02	596	291	681	370	642	287	465	217
aroC	2.06	1.07E-02	747	362	670	386	822	378	749	322
mreB	2.06	1.74E-02	783	377	676	407	919	378	754	346
rpsF	2.06	6.42E-03	3039	1474	3502	1629	3001	1516	2616	1277
dcm	2.06	9.05E-03	391	183	353	205	556	201	264	144
ascF	2.05	1.06E-02	4321	2105	4593	2433	5000	2214	3371	1668

ptsN	2.05	4.27E-03	1324	641	1568	799	1477	635	926	490
plsX	2.05	1.15E-02	190	96	217	128	166	95	186	64
insL	2.05	2.43E-04	194	94	315	148	117	54	149	80
prmA	2.05	6.02E-05	406	195	311	189	410	180	496	216
trmH	2.05	1.47E-04	203	98	232	117	237	104	139	73
gshB	2.04	1.16E-02	494	242	474	215	577	278	431	232
talB	2.04	6.37E-03	2120	1029	2551	1151	2195	1073	1614	863
fepB	2.03	9.90E-03	771	372	1106	483	656	353	553	281
icd	2.03	7.79E-07	11633	5726	13967	6958	11714	5461	9219	4758
ybiV	2.03	2.05E-02	261	126	450	210	192	94	140	74
ybaW	2.02	4.34E-02	103	55	105	80	98	55	106	30
yjcD	2.02	8.31E-04	280	142	276	179	300	148	265	100
purM	2.02	5.91E-03	2500	1224	2411	1147	3115	1432	1975	1095
tufB	2.00	8.28E-07	13817	6986	15324	8752	13250	6465	12879	5740
mltC	2.00	3.66E-04	426	213	361	170	519	246	400	225
rne	2.00	9.48E-03	1321	662	1181	671	1573	770	1208	544
yjhU	1.99	2.61E-02	287	140	211	134	273	148	377	138
asnS	1.99	6.80E-03	2226	1101	2591	1226	2497	1176	1589	900
prs	1.99	2.94E-05	2614	1309	3151	1518	2585	1329	2106	1080
dapF	1.99	3.46E-02	582	294	548	341	659	304	538	238
yfeZ	1.99	2.20E-02	196	99	263	136	139	81	187	79
yahL	1.99	3.18E-01	146	103	262	196			30	10
ydfG	1.98	4.34E-04	337	170	355	195	350	170	305	147
efeB	1.98	1.54E-02	464	227	651	261	398	236	342	185
pcnB	1.98	6.28E-03	641	325	751	394	660	324	513	256
rplC	1.98	1.83E-05	7953	3947	10228	4448	7572	3998	6059	3394
recF	1.98	3.40E-02	237	122	280	163	228	107	204	95
rplD	1.98	1.30E-02	12366	6120	16578	7269	11317	5930	9205	5163
ydfK	1.98	6.87E-03	1066	539	1210	587	1034	573	954	458
yjiA	1.98	6.08E-05	327	165	361	170	321	171	298	154
yeaP	1.98	3.73E-06	511	255	642	289	498	251	394	225
frmB	1.97	3.69E-02	346	184	335	266	355	132	347	153
pdxB	1.97	1.89E-02	386	196	441	232	357	181	359	175
oppA	1.97	9.25E-03	3682	1798	4378	1730	4432	2130	2235	1536
rplW	1.97	1.39E-02	7520	3816	6714	3734	8418	4159	7427	3554
accC	1.96	1.30E-06	1119	571	1197	629	1192	602	970	481
thrA	1.96	1.04E-02	1842	965	2072	1424	2251	877	1203	594
lysA	1.96	1.72E-02	370	186	319	196	482	203	309	159
rplB	1.96	1.31E-02	9292	4608	13754	5993	7198	4220	6925	3613
murC	1.95	1.66E-02	374	192	375	217	391	175	357	185
insC	1.95	1.73E-02	689	357	613	345	797	444	658	282
pepQ	1.95	2.59E-04	701	357	655	357	857	394	589	319
ygeI	1.95	2.21E-02	171	86	116	64	149	79	247	114

yciZ	1.94	1.01E-02	604	315	631	382	661	318	521	243
betT	1.94	1.58E-04	268	140	268	147	273	160	264	112
galR	1.94	2.32E-02	182	91	235	113	211	96	101	63
aspS	1.94	1.50E-02	871	437	1508	708	601	344	504	258
livM	1.94	8.24E-03	488	254	577	319	483	242	406	201
panB	1.94	1.69E-02	512	267	502	302	528	279	506	219
murB	1.93	1.60E-02	422	220	524	286	400	211	342	163
yibN	1.93	1.71E-02	592	310	581	344	649	342	547	243
gmk	1.93	9.55E-03	1222	637	1203	706	1402	695	1061	509
rpsP	1.93	1.58E-02	7060	3662	8148	4302	7279	3584	5752	3101
glgP	1.92	2.03E-02	397	200	566	250	376	201	249	149
cspF	1.92	1.73E-02	663	351	844	451	613	364	531	236
ribD	1.91	2.01E-05	435	227	535	267	417	226	355	188
xseB	1.91	1.50E-04	632	333	624	370	703	361	568	268
yebR	1.91	9.51E-03	1125	591	1263	701	1184	574	928	497
ilvH	1.91	1.83E-02	762	396	978	474	603	394	704	321
rpsB	1.90	5.40E-05	5419	2826	6572	3075	5151	2815	4534	2589
hyfG	1.89	4.67E-02	132	69	119	62	161	70	116	75
tolB	1.89	1.87E-02	1057	556	1185	585	1127	571	858	511
mioC	1.88	4.47E-05	879	466	980	546	965	464	694	387
gltX	1.88	1.97E-02	1236	656	1150	726	1486	709	1072	533
ybjX	1.88	1.76E-02	390	202	554	247	328	198	288	159
mdh	1.88	1.12E-02	2697	1425	2623	1504	3222	1544	2247	1228
priB	1.88	1.30E-02	4107	2143	5276	2275	3824	2233	3222	1922
rffH	1.88	1.94E-01	137	84	251	159	22	10		
yncG	1.88	8.53E-04	465	250	566	330	468	224	361	196
cydD	1.88	3.08E-02	228	119	197	113	184	113	303	129
yiiX	1.87	2.23E-02	218	119	244	152	212	120	197	85
rpsA	1.87	2.23E-02	4800	2500	7059	3292	3942	2337	3400	1870
rlmL	1.87	2.62E-02	513	266	358	252	660	290	521	256
ileS	1.87	2.02E-02	778	415	803	471	871	416	661	359
slmA	1.87	1.94E-02	314	169	302	192	289	178	351	137
gltB	1.87	1.16E-02	2095	1129	2265	1409	2325	1045	1696	933
zwf	1.86	3.91E-05	605	326	607	378	680	335	528	266
ninE	1.86	2.00E-01	260	127	23	26	722	344	36	10
tpx	1.86	1.14E-02	840	453	768	507	910	464	842	388
yjjX	1.86	3.75E-02	181	97	150	106	232	108	162	77
ymjA	1.85	2.17E-03	237	124	250	128	297	129	163	115
ulaE	1.85	2.81E-02	299	150	126	78	581	267	191	106
rumA	1.85	6.87E-02	84	45	63	57	115	40	74	37
gadC	1.85	3.20E-02	183	101	145	67	223	116	182	118
ycdX	1.85	1.12E-04	458	249	461	249	436	276	477	220
livF	1.85	1.04E-04	1219	667	1433	840	1188	643	1036	517

fabH	1.85	1.30E-02	1055	569	1026	623	1264	614	875	468
rhsC	1.84	1.89E-02	204	112	239	135	191	108	183	91
dam	1.84	4.08E-02	105	57	77	65	128	62	108	43
modC	1.84	3.02E-03	1266	676	1730	717	1127	784	941	526
glyA	1.84	4.34E-04	4587	2464	5492	2632	4836	2665	3434	2094
rplN	1.84	1.77E-04	17970	9613	22208	11676	19363	9171	12340	7991
emtA	1.83	6.32E-02	296	167	457	276	210	119	221	104
nlpC	1.83	2.54E-02	432	240	486	285	421	260	388	175
yceI	1.83	3.01E-02	208	112	162	109	228	115	234	112
yqjB	1.83	2.01E-04	168	93	170	107	172	100	162	73
rfe	1.83	1.68E-02	649	360	719	425	618	377	609	277
rffA	1.82	2.96E-03	146	80	160	75	139	86	138	79
acrA	1.82	9.21E-05	914	502	1123	655	987	476	631	374
setC	1.82	2.67E-02	486	271	518	282	468	321	471	211
ypjK	1.81	3.27E-01	244	291	587	809	44	10	101	54
lon	1.81	1.07E-02	541	306	574	396	592	303	457	218
yidC	1.81	3.89E-04	890	511	1258	799	727	396	687	337
yheM	1.80	2.98E-02	889	455	2027	988	289	204	350	173
add	1.80	1.58E-03	186	101	163	99	244	103	151	100
yliG	1.80	3.23E-02	238	131	204	131	226	137	283	123
pdxJ	1.80	1.63E-03	399	227	456	301	371	207	370	171
nhaB	1.80	1.12E-01	95	53	85	66	75	53	126	39
ubiD	1.80	4.90E-03	447	250	431	284	413	249	496	216
metC	1.80	3.87E-01	1491	413	3915	679	283	310	275	250
pntA	1.79	1.25E-02	2731	1554	3039	1977	2998	1631	2156	1054
rihB	1.79	1.39E-01	317	184	308	246	278	188	367	118
dipZ	1.79	9.48E-03	187	102	151	92	254	112	157	101
trmI	1.78	1.24E-02	667	351	1315	644	385	210	302	199
yheN	1.77	2.11E-02	445	252	416	249	472	283	448	224
cyaY	1.77	6.28E-02	194	108	142	111	200	126	242	89
sufB	1.76	2.77E-04	1577	882	2079	1027	1394	893	1259	726
tatC	1.76	7.54E-02	279	158	306	190	201	166	331	117
helD	1.76	2.89E-02	169	93	216	114	189	83	102	80
yhhS	1.76	2.06E-01	123	62	78	75	218	59	73	52
yffS	1.75	1.85E-01	257	154	257	203	212	173	301	87
typA	1.75	1.40E-02	1806	1020	1953	1193	2210	1074	1255	794
yeiR	1.75	3.85E-02	269	152	234	155	237	142	336	159
carB	1.74	3.00E-02	1395	809	1635	860	1231	961	1318	607
hemG	1.73	4.68E-02	523	284	807	329	337	283	426	241
ybhP	1.73	1.21E-01	186	122	201	209	228	102	130	54
rfaD	1.72	5.12E-02	525	311	605	380	486	334	485	219
surE	1.72	2.22E-04	379	219	327	217	384	235	425	206
groS	1.72	1.65E-02	1858	1067	2386	1159	1749	1168	1439	873

ribE	1.71	5.37E-04	641	378	648	440	611	382	662	312
ybfE	1.70	6.50E-02	167	98	128	118	205	89	169	86
nrfF	1.70	3.74E-02	102	57	121	64	129	61	56	47
yagL	1.69	4.08E-01	280	120	790	320	28	10	22	31
yaiO	1.69	4.45E-01	380	147	1063	394	10	10	67	38
metH	1.69	1.85E-03	477	282	558	265	452	303	422	278
phnI	1.68	1.74E-01	142	96	181	174	138	60	108	54
fimC	1.68	3.81E-02	794	475	672	491	1017	591	693	344
gsiD	1.68	4.17E-02	286	170	287	203	237	160	333	147
csdA	1.67	5.08E-03	265	168	354	266	234	137	207	101
yffO	1.66	6.41E-02	907	610	1385	1171	902	413	432	246
mdtC	1.65	2.07E-01	162	90	272	93	98	111	117	66
grxB	1.65	4.97E-02	536	326	469	352	577	348	562	277
dgsA	1.63	3.64E-02	319	181	228	170	201	145	528	229
hslO	1.62	6.37E-02	257	157	204	135	258	187	308	150
ksgA	1.62	5.24E-02	397	242	467	328	477	219	248	180
mdtQ	1.62	6.19E-01	196	208	270	384			122	33
livH	1.61	6.12E-02	460	280	465	321	342	268	572	252
yciK	1.61	1.30E-01	232	141	161	158	248	152	286	113
orn	1.61	8.22E-02	172	97	65	61	275	148	176	84
skp	1.60	6.41E-02	2616	1550	3698	1826	2628	1559	1521	1266
hyfR	1.60	3.70E-02	136	87	104	77	155	121	149	63
melR	1.59	4.37E-02	195	120	163	127	254	121	168	111
mrdB	1.58	3.54E-01	119	87	138	41	107	67	112	152
tdk	1.57	7.37E-03	183	121	227	174	154	104	169	84
yraJ	1.54	1.92E-01	169	114	130	123	196	152	181	68
glnE	1.54	4.96E-02	368	238	334	254	335	245	436	216
rpsE	1.54	1.29E-02	8500	5038	15458	7558	5271	4046	4769	3511
oxc	1.54	1.28E-01	330	219	234	233	465	296	291	127
yjdF	1.53	1.54E-01	291	210	307	128	472	432	95	69
ybjI	1.53	2.05E-01	191	123	177	83	127	148	270	137
selA	1.53	1.41E-01	188	124	152	124	187	156	226	92
hemX	1.53	1.04E-02	418	277	490	376	317	233	447	223
yjjV	1.52	2.56E-01	145	91	197	99	155	70	84	105
yfgA	1.52	1.01E-01	699	453	530	472	709	470	858	416
flgI	1.51	1.11E-01	453	328	275	127	521	412	563	445
yfcA	1.50	1.11E-01	327	241	186	83	460	348	336	291
ybhU	1.49	1.53E-01	101	64	186	118	33	36	85	37
yfcN	1.48	9.01E-02	265	177	244	204	225	178	325	151
malZ	1.48	2.06E-01	119	75	128	85	166	77	64	64
yegW	1.47	2.00E-01	139	100	127	102	147	139	143	58
yfhG	1.47	3.59E-01	525	292	1026	328	292	298	258	249
nhoA	1.47	3.98E-01	90	83	189	195	27	10	54	45

insF	1.45	7.00E-01	1264	352	3380	408	292	428	120	221
yfaY	1.45	1.96E-01	131	91	151	69	130	93	111	113
clpB	1.43	1.56E-01	295	205	228	233	305	208	350	173
mgtA	1.42	2.88E-01	145	85	263	117	123	75	50	64
intA	1.42	2.20E-01	188	134	238	193	126	111	200	98
dgoT	1.41	4.72E-01	130	100	139	89	100	158	150	52
citD	1.41	5.30E-01	95	67	10	10	254	180	20	10
mdoD	1.41	2.86E-01	754	386	1845	760	226	246	191	152
cynX	1.38	3.33E-01	86	63	91	61	73	83	94	46
ybbJ	1.34	6.88E-01	28	27			32	44	25	10
ybjQ	1.32	3.16E-01	257	192	193	210	257	216	320	151
insA	1.29	5.60E-01	59	42	47	42	35	38	94	44
yaeR	1.27	4.91E-01	521	309	1138	461	221	247	203	218
ybgA	1.27	6.53E-01	776	302	2114	647	78	134	136	125
uhpB	1.27	5.53E-01	193	107	443	168	55	69	83	84
yejH	1.27	4.31E-01	250	181	280	228	352	172	117	144
ybaV	1.25	5.09E-01	40	38	27	13	51	37	43	64
ynaI	1.23	6.13E-01	528	244	1452	581	77	70	56	82
ddpB	1.23	5.76E-01	101	68	26	39	112	89	166	76
poxA	1.22	5.99E-01	786	434	1983	906	243	195	132	200
ppdB	1.18	7.43E-01	129	140	56	66	128	44	202	310
hchA	1.16	6.28E-01	109	97	133	66	97	97	98	127
ybbM	1.14	7.04E-01	108	99	105	48	165	174	54	74
tfaD	1.14	7.06E-01	119	101	84	81	93	136	181	87
yfhD	1.11	8.36E-01	401	257	852	376	262	193	91	204
ypjF	1.11	9.18E-01	167	86	200	127	13	64	289	65
yohK	1.10	8.64E-01	111	78	44	64	60	96	229	75
ynaA	1.07	9.31E-01	62	103	103	197			22	10
yeaL	1.03	9.70E-01	75	55	5	16	63	23	157	126
menE	1.03	9.66E-01	795	381	2102	660	177	288	108	195
rcnA	0.95	9.15E-01	351	213	39	93	157	185	857	361
yoeA	0.93	9.19E-01	275	254	529	157	145	299	152	306
yjeP	0.91	8.52E-01	58	87	87	121	20	10	68	130
wcaM	0.91	8.42E-01	104	68	258	113	28	41	25	52
ynbB	0.90	8.40E-01	206	164	27	79	284	165	307	247
napF	0.90	8.13E-01	49	44	39	36	16	41	93	53
yqhH	0.89	7.94E-01	918	680	2169	1028	326	533	259	478
yoeD	0.86	7.48E-01	645	403	1745	899	64	147	126	164
recX	0.86	7.27E-01	361	331	370	408	627	384	87	200
yedK	0.82	5.75E-01	95	106	112	103	128	105	46	109
artM	0.82	6.82E-01	979	653	2586	1289	175	368	175	300
ttdA	0.81	6.39E-01	136	147	31	73	183	258	196	110
ycjY	0.80	7.70E-01	112	80	216	144			7	16

phnJ	0.80	5.80E-01	520	406	1457	1046	46	101	58	70
yihG	0.80	6.22E-01	75	81	121	92	77	85	29	66
glxR	0.80	7.19E-01	115	109	39	59	52	160	255	106
bhsA	0.79	3.30E-01	118	146	107	103	170	177	78	157
yehU	0.78	6.97E-01	295	319	204	554	87	167	595	237
yagX	0.78	5.67E-01	136	142	84	107	269	192	56	128
glxK	0.77	4.34E-01	65	74	68	74	98	82	28	66
iraP	0.77	5.91E-01	1649	1280	4131	2183	463	970	352	689
glvC	0.77	4.94E-01	344	417	172	355	456	504	403	391
yfbV	0.76	5.08E-01	502	585	676	754	638	538	192	462
insA	0.76	7.52E-01	151	113	392	221	49	28	13	91
yjeF	0.75	4.32E-01	106	129	154	148	55	123	109	117
yhdY	0.75	4.75E-01	133	162	97	195	84	148	218	143
yhjK	0.75	5.73E-01	72	85	74	71	113	81	29	102
yfbB	0.75	4.52E-01	75	101	88	87	58	123	81	91
yfcO	0.74	4.58E-01	326	425	204	226	550	579	225	470
yfcI	0.74	4.02E-01	178	205	47	106	266	263	222	248
ulaR	0.73	5.65E-01	200	172	498	310	65	75	37	132
rtcB	0.73	5.20E-01	142	181	250	272	62	173	114	97
gudD	0.72	4.25E-01	153	171	280	208	50	141	130	165
yjcS	0.72	3.14E-01	86	98	18	43	151	156	89	94
nac	0.72	3.19E-01	534	509	1319	996	130	281	151	249
hcaT	0.72	3.23E-01	77	106	68	145	54	82	110	91
ybjO	0.71	3.41E-01	69	98	75	56	40	96	91	140
ybaT	0.71	3.93E-01	67	94	72	78	52	108	79	94
citC	0.71	6.77E-01	736	500	1453	945			18	55
glcA	0.71	5.24E-01	89	101	37	52	36	83	193	169
ulaD	0.70	4.47E-01	32	43	18	19	59	57	19	54
ycgK	0.70	2.31E-01	121	158	131	171	170	173	61	132
ygbN	0.70	3.73E-01	178	261	202	180	171	347	161	257
yjeO	0.70	2.02E-01	112	153	65	81	193	209	79	169
mdtK	0.70	2.59E-01	183	211	381	341	62	146	106	146
yccJ	0.70	4.84E-01	235	274	384	315	251	280	71	228
smf	0.70	3.99E-01	331	438	271	416	525	505	196	392
yjeH	0.69	4.60E-01	114	156	115	131	163	181	65	156
elaD	0.69	5.75E-01	55	59	7	21	135	132	24	25
yhcD	0.68	1.16E-01	170	234	123	165	276	302	111	235
proA	0.68	2.92E-01	402	570	297	605	556	598	355	505
mdtP	0.68	2.34E-01	102	131	57	112	182	146	67	134
argA	0.68	4.05E-01	237	321	165	350	159	338	386	275
rtcR	0.68	1.11E-01	128	191	136	135	154	250	93	188
menC	0.67	3.22E-01	105	151	139	163	102	140	75	150
insB	0.67	3.60E-01	311	472	304	614	347	398	283	404

yohO	0.67	3.25E-01	104	147	157	150	78	171	77	121
ebgA	0.67	1.90E-01	595	876	783	814	519	1053	485	760
ais	0.66	3.16E-01	154	188	44	120	281	257	139	188
cobU	0.66	2.59E-01	204	285	297	328	215	285	102	241
yhaI	0.66	3.20E-01	111	162	92	103	174	222	66	162
rutR	0.66	3.59E-01	191	287	102	151	174	411	296	298
tfaE	0.65	2.51E-01	294	441	201	420	376	531	306	371
yihQ	0.65	3.94E-01	80	126	20	29	98	200	123	149
proY	0.65	3.48E-01	322	456	273	445	511	544	182	379
napG	0.65	3.35E-01	89	132	69	103	68	143	131	150
ybfD	0.65	1.70E-01	82	120	115	144	77	106	53	111
fhlA	0.65	2.52E-01	127	191	140	207	156	181	85	184
yhjY	0.65	1.55E-01	76	117	83	104	58	131	87	115
yedZ	0.64	3.77E-01	112	149	203	200	46	118	87	128
rimL	0.64	2.92E-01	99	143	112	150	131	143	54	137
eptB	0.64	3.26E-01	442	720	501	1042	522	713	304	405
yegI	0.64	4.91E-01	38	64	30	39	36	48	47	105
aas	0.64	3.36E-01	164	294	45	53	282	566	164	264
amiD	0.64	2.36E-01	129	202	149	208	110	222	129	175
insF	0.64	2.49E-01	183	267	105	230	164	316	280	254
yeiN	0.64	3.52E-01	114	193	60	71	118	287	165	222
dsdC	0.63	2.21E-01	219	339	162	329	302	413	194	274
spy	0.63	3.42E-01	153	236	125	254	206	248	128	206
yafE	0.63	3.83E-01	137	197	226	215	92	214	92	164
yeaV	0.63	2.52E-01	192	300	204	250	228	339	145	312
ymdC	0.63	3.55E-01	121	197	128	136	104	243	132	212
eutD	0.63	2.28E-01	79	93	39	65	175	144	24	70
renD	0.63	2.31E-01	762	1190	842	1759	1176	1370	269	443
uvrY	0.63	8.14E-02	1187	1782	607	1290	1716	2328	1239	1727
sgbE	0.63	2.25E-01	536	813	343	755	851	1092	415	591
trg	0.63	1.49E-01	146	184	320	314	54	118	65	121
intG	0.63	1.06E-01	112	182	120	141	100	226	117	178
rhsA	0.63	1.26E-01	96	156	102	113	92	188	93	166
yiaD	0.62	2.74E-01	127	190	75	151	212	270	94	149
sfmA	0.62	2.33E-01	64	86	30	43	134	127	29	89
ytjA	0.62	3.16E-01	192	323	134	158	176	396	265	415
pinR	0.62	3.34E-01	156	257	127	158	128	351	212	262
yffR	0.62	3.13E-01	222	350	206	275	154	361	306	414
rfaZ	0.61	5.49E-02	114	183	138	165	119	193	86	190
wza	0.61	1.93E-01	108	169	131	143	130	216	62	146
yjbI	0.61	2.08E-01	135	218	104	153	195	281	108	219
ydbD	0.61	2.19E-01	145	233	189	227	127	222	121	249
yfeR	0.61	3.46E-01	93	143	47	103	93	183	140	142

agaC	0.61	3.01E-01	99	142	145	186	36	127	116	111
yeeS	0.61	6.88E-02	115	184	139	182	131	214	77	157
yqjF	0.61	2.91E-01	775	1244	1232	1938	660	890	434	905
yghA	0.61	4.97E-01	267	233	674	292	56	215	72	192
araD	0.61	2.73E-01	90	149	74	136	89	192	108	118
ygdI	0.61	2.10E-01	166	251	81	209	225	293	192	252
sdaA	0.61	8.33E-02	344	525	437	581	429	589	166	404
gspI	0.61	2.84E-01	199	293	61	128	352	414	184	335
mokA	0.60	2.11E-01	564	919	455	978	802	1126	436	652
csgD	0.60	2.44E-01	96	158	113	135	103	165	74	174
tnaA	0.60	2.70E-01	183	244	342	281	150	254	59	197
yfdS	0.60	1.64E-01	335	539	206	419	507	730	292	468
tdcG	0.60	2.39E-01	102	156	40	88	176	239	90	141
kbaZ	0.59	3.02E-01	419	604	248	561	819	873	191	377
yafW	0.59	2.47E-01	139	219	215	306	136	195	67	157
yadC	0.59	2.61E-01	123	200	100	184	96	195	174	222
ybjJ	0.59	3.82E-01	83	142	108	92	78	151	63	184
dctR	0.59	3.54E-01	75	124	89	146	87	124	50	101
yidI	0.59	3.35E-01	210	293	433	438	99	260	99	182
glnK	0.59	2.15E-01	666	1101	822	1136	750	1290	425	877
pphA	0.59	2.84E-01	232	389	306	460	215	344	176	362
ygeX	0.59	1.80E-01	449	722	725	1003	373	636	251	527
yahI	0.59	1.82E-01	189	327	212	431	205	306	149	245
yibG	0.58	1.05E-01	83	135	54	93	138	198	56	115
ydhW	0.58	2.00E-01	133	226	130	263	169	229	101	186
kdpD	0.58	2.24E-01	140	209	126	214	238	264	56	149
eutP	0.58	8.31E-02	100	171	100	154	116	188	83	170
ypaB	0.58	1.69E-01	753	1330	927	1919	884	1339	450	733
yibJ	0.58	6.72E-02	234	393	237	352	303	447	162	379
arsC	0.58	2.68E-01	400	669	628	959	264	544	309	505
arpA	0.58	2.52E-01	214	311	158	474	68	165	415	295
hokE	0.58	1.65E-01	674	1171	642	1134	623	1337	757	1043
pyrE	0.58	2.67E-01	407	717	381	511	352	823	488	817
phoA	0.58	2.97E-01	98	158	49	92	167	167	77	215
dinQ	0.57	3.25E-01	115	191	176	218	83	149	87	206
yadD	0.57	1.98E-01	117	200	146	184	110	210	95	208
tnaB	0.57	1.54E-01	280	491	313	586	239	481	287	405
csgG	0.57	3.42E-01	92	157	100	120	118	190	58	162
yeeL	0.57	2.14E-01	83	157	68	77	89	213	91	180
yghG	0.57	3.16E-01	108	165	34	79	92	233	199	183
gatD	0.57	5.61E-02	332	579	304	609	431	676	261	451
ecnB	0.57	1.62E-01	228	396	210	344	287	443	187	400
narW	0.57	5.14E-02	335	581	242	491	482	796	281	457

aldB	0.57	1.47E-01	3553	6514	5537	11120	3312	5378	1812	3045
ycbJ	0.57	1.61E-01	411	721	334	808	435	713	464	642
csiE	0.57	2.63E-01	643	1056	590	1122	1085	1519	254	525
paaJ	0.57	3.21E-02	382	637	573	852	397	662	178	396
ykgL	0.57	2.56E-01	118	143	300	320	16	55	37	55
yjfI	0.57	3.35E-01	76	134	77	112	92	154	60	137
rcsC	0.56	2.88E-01	159	271	232	328	138	255	108	231
ogt	0.56	3.18E-01	101	164	52	163	154	187	96	140
agaS	0.56	2.51E-01	58	100	44	90	51	96	78	113
ykgM	0.56	9.34E-02	96	175	82	108	82	212	125	205
ybhQ	0.56	2.74E-01	343	558	512	663	366	640	150	370
yciN	0.56	4.97E-02	195	339	259	396	193	338	133	283
yfjX	0.56	7.73E-02	127	220	53	103	182	262	146	295
zraP	0.56	6.63E-02	262	441	285	538	369	475	131	309
wcaE	0.56	3.30E-01	99	169	76	89	54	206	166	214
ydaN	0.56	2.57E-01	399	681	705	1077	261	559	232	408
lsrG	0.56	7.69E-02	88	157	54	90	136	223	75	158
yphG	0.56	5.28E-02	368	667	432	773	359	734	313	496
frvB	0.56	7.98E-02	143	241	257	381	104	218	67	125
iap	0.56	2.09E-01	97	177	105	140	92	183	96	209
wcaI	0.55	2.45E-01	835	1505	751	1223	1064	1868	691	1423
yihY	0.55	2.57E-01	69	112	75	116	100	107	32	113
aroM	0.55	4.62E-02	213	374	250	376	246	386	143	360
fabA	0.55	4.35E-02	1625	2884	2223	3672	1564	2754	1089	2225
yfeS	0.55	2.55E-01	564	966	1022	1451	456	1064	214	384
yaaW	0.55	8.52E-02	150	272	120	197	204	348	126	272
ddpX	0.55	1.81E-01	103	184	70	157	123	216	117	178
gspD	0.55	7.38E-02	172	303	242	335	160	335	114	240
yjeT	0.55	4.13E-02	517	929	782	1326	369	738	401	722
ycjU	0.55	2.76E-01	236	366	69	200	238	488	400	411
uspB	0.55	6.60E-02	122	226	132	174	102	285	132	219
metJ	0.55	1.48E-01	787	1488	991	2073	694	1428	676	963
yeaG	0.55	3.63E-02	367	670	449	768	310	694	341	548
yagA	0.54	1.45E-01	352	610	530	820	360	631	167	380
ygcE	0.54	1.55E-01	455	833	681	1136	372	857	313	504
yfcV	0.54	4.46E-02	196	354	211	341	229	397	147	326
cmtB	0.54	8.43E-02	100	165	84	126	168	223	47	145
ysaA	0.54	2.92E-01	134	240	207	357	120	213	74	151
ydfV	0.54	6.87E-02	282	473	296	602	438	537	112	280
ydhY	0.54	1.70E-01	158	286	212	336	120	243	142	277
mdtG	0.54	2.92E-01	302	439	625	668	199	420	81	228
rplT	0.54	1.35E-01	599	1126	651	1394	609	1116	537	869
yaaJ	0.54	2.58E-01	187	342	177	369	258	424	125	233

lacZ	0.54	1.52E-01	316	584	309	570	347	596	291	584
arnE	0.54	1.42E-01	61	108	52	73	95	110	35	141
frlD	0.54	1.64E-01	170	308	178	383	223	294	109	245
hepA	0.54	1.56E-01	540	1009	573	1172	582	1006	466	850
astA	0.54	1.55E-01	981	1791	1457	2381	930	1932	555	1058
mhpC	0.54	2.39E-01	609	1063	1036	1558	456	881	336	751
ygaU	0.54	6.05E-02	95	178	106	196	82	169	98	168
eutC	0.53	2.50E-01	82	138	39	145	122	138	86	131
ydjI	0.53	3.60E-02	428	805	362	716	489	980	434	718
caiF	0.53	2.99E-01	194	316	245	291	269	419	67	239
yegD	0.53	1.80E-01	139	255	178	204	139	280	99	281
yneN	0.53	1.60E-01	190	347	150	359	259	381	161	302
yjdA	0.53	1.41E-01	235	447	274	537	204	465	227	338
ygaC	0.53	1.52E-01	497	921	496	1105	668	1012	326	644
ytfK	0.53	1.49E-01	477	906	550	1051	461	966	420	701
tisB	0.53	4.03E-01	170	130	468	226	20	64	23	100
yghF	0.53	2.64E-01	104	202	74	121	140	283	98	203
glpF	0.53	1.04E-01	104	171	61	140	191	191	61	183
mpaA	0.53	3.26E-01	102	178	116	131	144	207	46	195
yfaT	0.52	3.05E-01	190	286	412	473	97	233	60	150
sirA	0.52	1.74E-01	221	412	181	438	305	467	178	331
yhdL	0.52	2.49E-01	137	262	126	237	157	287	128	262
ydfA	0.52	1.41E-01	377	715	425	861	455	742	253	543
gabT	0.52	1.34E-01	414	818	418	1003	523	969	302	483
arnA	0.52	2.14E-01	88	151	74	104	150	194	40	154
glcE	0.52	3.84E-02	163	301	244	367	115	281	132	256
ypjC	0.52	1.37E-01	239	452	276	410	173	464	269	483
yhcC	0.52	1.44E-01	395	758	506	905	352	739	327	629
yehB	0.52	1.72E-01	66	139	48	63	85	199	67	156
ygaM	0.51	2.26E-01	527	1011	402	927	668	1182	511	924
rpoS	0.51	2.22E-01	5414	10605	5530	12064	5979	10756	4733	8994
gudP	0.51	2.22E-01	618	1198	734	1334	678	1324	443	936
potE	0.51	3.17E-01	124	200	105	170	224	228	44	200
ybdJ	0.51	1.73E-01	94	176	136	198	89	179	59	151
hyfI	0.51	1.25E-01	264	515	309	551	263	535	220	459
prpD	0.50	2.25E-01	141	233	277	323	61	213	84	162
wcaA	0.50	1.29E-01	300	594	342	641	310	596	250	546
yfaE	0.50	2.17E-01	1283	2515	2015	3906	1109	1995	725	1644
eamB	0.50	2.23E-01	392	790	303	561	388	1033	484	777
yehP	0.50	1.44E-01	359	725	479	953	355	743	242	477
yddM	0.50	3.08E-02	295	596	318	678	314	592	253	517
mqsR	0.49	1.28E-01	431	842	629	1019	334	828	330	678
phnD	0.49	1.16E-01	945	1952	915	2171	987	2127	935	1557

ydaW	0.48	2.79E-01	146	242	286	322	52	199	99	204
yjbJ	0.48	2.13E-01	914	1845	767	1916	1266	2002	709	1617
dinG	0.48	2.18E-01	192	399	200	445	202	395	173	359
tfaX	0.48	1.44E-01	57	94	79	143	12	55	81	84
yjaZ	0.48	2.27E-01	183	373	220	424	200	335	128	360
glpQ	0.48	1.16E-01	247	515	224	519	313	607	205	421
arnB	0.48	2.18E-01	402	734	837	1293	183	428	187	479
paaK	0.48	1.23E-01	319	668	329	714	362	690	268	600
ydhI	0.48	4.30E-01	56	117			58	108	53	126
iscR	0.47	1.04E-02	430	896	525	866	365	962	400	861
yhaM	0.47	2.10E-01	825	1717	594	1612	1022	2095	859	1445
iscU	0.47	1.19E-01	692	1422	1064	1972	559	1328	453	966
qor	0.47	2.28E-01	152	312	202	343	149	334	104	259
mtfA	0.47	1.11E-01	407	861	486	963	408	871	329	750
yeaH	0.47	3.08E-02	147	310	160	322	164	320	116	288
ftnB	0.47	1.40E-01	99	205	115	213	121	227	63	176
yjjM	0.47	2.52E-01	155	169	408	291	29	96	30	121
ydaT	0.47	1.13E-01	217	439	93	235	231	565	328	519
fumA	0.47	1.18E-01	642	1373	805	1726	481	1330	640	1062
yehR	0.47	4.80E-01	57	86	108	155			6	18
chpS	0.46	2.18E-01	140	301	141	286	161	326	118	291
yhbQ	0.46	8.29E-02	85	162	64	164	148	171	43	150
yciG	0.46	1.08E-01	310	669	357	700	323	691	251	615
ycfK	0.46	1.43E-01	86	187	50	105	91	226	118	228
gatC	0.46	3.32E-02	130	282	151	266	131	300	109	280
uspF	0.46	2.26E-02	188	409	230	446	174	411	161	369
ygfJ	0.46	2.20E-01	235	488	358	741	249	445	98	278
dpiA	0.45	3.60E-01	89	108	43	76	212	136	12	113
ymdF	0.45	1.08E-01	232	505	245	566	273	501	179	449
epd	0.45	2.05E-01	216	472	242	508	161	473	246	434
argC	0.45	1.13E-01	201	436	258	486	202	431	143	391
yaiV	0.45	3.57E-02	89	195	101	170	107	251	58	165
ydcS	0.45	1.06E-01	283	604	443	811	222	540	183	462
ygiT	0.45	1.25E-01	164	359	163	407	216	372	114	297
napA	0.45	1.34E-01	118	249	147	305	148	254	60	188
yieF	0.44	2.01E-01	303	646	495	852	203	622	212	464
ydcV	0.44	1.25E-01	109	222	41	137	163	276	124	254
yjcH	0.44	1.96E-01	265	606	258	678	289	638	248	502
ygfQ	0.44	2.33E-01	131	256	70	223	250	358	73	188
yfhL	0.44	1.97E-01	216	492	236	480	248	586	164	411
yobH	0.43	2.10E-01	195	399	77	305	258	459	250	433
yaiS	0.43	2.24E-01	123	279	127	225	178	424	65	190
ddpF	0.43	2.13E-01	94	220	81	175	92	258	110	225

argK	0.43	1.25E-01	98	212	45	116	166	293	83	227
ygdT	0.43	2.69E-01	87	175	156	176	60	164	45	183
tdcE	0.43	2.05E-01	274	520	29	110	360	636	434	814
yneM	0.43	9.62E-02	546	1260	648	1301	564	1391	428	1088
nanT	0.43	2.24E-01	338	364	948	849	27	118	39	125
yiiU	0.43	1.11E-01	864	1818	1524	2498	656	1734	414	1222
yhhA	0.42	2.24E-01	72	165	48	142	69	187	100	167
yqeJ	0.42	1.99E-01	214	486	320	655	218	520	103	285
tisA	0.41	9.63E-02	342	809	420	960	385	798	221	668
ynjI	0.41	1.95E-01	148	351	153	307	100	333	192	413
tdcA	0.41	1.37E-01	114	225	35	162	213	323	94	191
ychH	0.41	1.84E-01	378	888	528	1043	366	919	239	702
treR	0.41	2.12E-01	107	255	135	300	117	262	69	202
yqfA	0.41	1.56E-02	416	1022	484	1178	366	1019	397	869
yceJ	0.41	2.09E-02	176	417	232	398	179	464	116	388
yjcZ	0.41	1.94E-01	152	366	126	274	225	490	106	333
yohC	0.40	1.86E-01	185	453	221	505	139	388	195	466
cspA	0.40	8.89E-02	8695	22032	11867	31946	7921	19001	6297	15150
yaiY	0.40	1.01E-01	148	369	179	417	121	349	146	340
treC	0.39	1.86E-01	52	110	75	108	16	113	66	108
glgS	0.39	2.01E-01	99	256	79	241	95	334	123	193
gabD	0.39	1.63E-02	175	452	186	540	152	428	187	386
yodD	0.39	2.14E-02	92	239	87	270	90	250	99	198
yqeC	0.38	1.78E-01	153	396	103	291	152	386	205	512
dmsB	0.38	2.76E-01	44	65	84	112	3	17		
yebV	0.38	1.73E-01	621	1624	529	1693	711	1801	622	1379
cspE	0.38	3.84E-03	2005	5282	2460	6600	1878	4980	1678	4265
yiaG	0.38	1.41E-02	251	661	283	767	262	662	208	556
setA	0.38	1.76E-01	326	845	317	925	245	838	417	772
pspE	0.37	2.06E-01	119	281	208	321	84	286	64	235
rmf	0.37	8.29E-02	22588	60965	32257	88067	19747	51381	15759	43447
phnC	0.37	2.79E-01	114	134	308	201	23	60	12	141
crp	0.37	8.40E-02	539	1472	479	1573	613	1625	526	1219
glcF	0.37	1.83E-01	113	301	80	287	114	327	146	288
ygcU	0.37	2.07E-01	68	181	55	158	97	233	51	152
yihO	0.36	8.36E-02	63	136	62	129	107	140	18	140
glcD	0.36	1.29E-02	325	843	515	1111	278	806	182	611
erpA	0.36	8.79E-02	330	910	233	832	392	1048	367	850
astC	0.35	1.41E-01	62	135	22	88	128	168	35	150
aroF	0.35	1.08E-02	2487	7037	3010	8213	2542	6990	1907	5907
ydeJ	0.35	1.26E-01	60	166	74	148	71	187	37	163
arrQ	0.34	1.04E-01	78	218	124	277	57	209	55	169
treB	0.34	1.43E-01	60	137	42	148	20	140	119	124

lrhA	0.32	1.68E-01	128	387	144	386	93	418	147	358
yfiD	0.32	7.55E-02	1962	6120	2144	6654	2065	6526	1676	5182
prpC	0.32	1.74E-01	76	226	38	155	108	284	83	239
ybaJ	0.31	1.18E-02	140	445	162	511	139	419	118	404
gltL	0.31	2.66E-01	38	51	83	72	1	20	31	61
bssR	0.31	7.56E-02	519	1733	600	2314	575	1731	382	1152
yjjQ	0.31	6.02E-02	43	89	104	128	9	50	17	90
yraH	0.30	3.42E-01	19	54			29	66	9	42
yjfO	0.30	9.37E-03	475	1584	357	1618	611	1800	458	1333
fabB	0.29	7.13E-02	576	1946	764	2226	572	2058	393	1554
glpT	0.28	9.47E-03	147	502	206	579	95	477	139	449
yfdT	0.27	1.05E-01	34	128	18	54	52	157	34	172
yhbO	0.27	2.71E-01	113	228	233	192	96	219	11	273
tyrA	0.27	6.92E-02	1043	3768	1343	4276	1083	3937	703	3092
asnA	0.27	7.08E-02	162	603	165	647	171	636	151	528
yciY	0.27	6.98E-02	283	1059	288	1107	286	1070	277	1001
pspD	0.25	1.63E-01	61	235	80	258	36	230	68	216
feaR	0.25	9.53E-02	51	168	72	159	65	140	15	205
sfsA	0.23	1.77E-03	152	652	179	688	113	670	166	599
ribB	0.22	7.38E-02	80	324	130	342	73	334	39	297
ypeC	0.21	1.47E-01	69	277	27	235	68	323	112	273
ulaF	0.21	1.71E-01	37	164	57	116	29	203	24	172
ycfJ	0.21	1.33E-01	240	1084	240	1233	333	1185	146	833
raiA	0.21	6.15E-02	1994	9302	2599	9853	1888	10116	1497	7937
nikE	0.20	6.61E-01	15	2743	-16	5474			46	12
pspB	0.19	1.38E-03	154	800	170	883	139	808	152	709
yrdE	0.19	3.24E-01	38	50	119	108	-11	31	7	11
elbA	0.16	no replicates	19	117					19	117
cpxP	0.15	5.53E-02	191	1239	254	1529	165	1199	154	988
fryB	0.15	2.29E-01	50	110	123	157	-8	83	36	89
glpA	0.14	7.34E-02	27	164	50	120	13	157	18	215
ybfB	0.13	4.98E-01	1411	790	2831	1568	-9	12		
htrE	0.12	5.19E-01	22	24			55	36	-10	13
tsgA	0.12	4.68E-01	197	155	406	300	-12	10		
mngB	0.12	4.85E-01	61	67	128	95	-5	39		
ydcH	0.10	4.85E-02	374	3531	505	4094	365	3754	253	2746
osmB	0.10	4.84E-02	107	1085	122	1074	97	1230	103	951
pspA	0.09	4.73E-02	153	1579	173	1704	191	1718	94	1316
yehK	0.09	1.30E-01	26	136	-33	68	24	149	87	189
prpB	0.09	1.25E-01	92	320	26	282	272	346	-23	330
frlB	0.08	3.63E-01	31	68	71	118	-10	18		
yjfN	0.07	3.59E-02	55	470	145	519	-25	508	44	383

Appendix II

Sequences obtained through ChIP-seq and used in MEME to predict YebK binding consensus. Sequence present in the direct strand are marked in *green* and in the complementary strand are marked in *red*. Summary of this data is described in Chapter 5.

Genome locus	Sequence
>de3-124401-124501	GTGTGTAAGTTTGCAATTCCGTTTGTTGTATTAATTTGTTTACATCAAAGA**AGTTTGAATTGTTACAA**AAAGACTTCCG TCAGATCAAGAATAATGGTATG
>de3-332351-332501	CACCGGCGTTAAAACAAAAGCCTAGATAAATACCACCAACAATCGATCCTACTAAA**ATGTTGTATTGTAACAG**TGGCCC GAAGATAAAAATAAAGAACGGCGCAAACATCACTAACATGCCGGTAATAATCCACAGCAGGTATTTGCGCAG
>de3-397551-397701	TTACGTAATCGATTGCGTTCACGTTTACACACATTCGGGCAGGGATTGTACTGACTTTCACCC**TTGTTGCAAACTTTTA C**TATCAATCTGGGTGATGTAAGCGGAGTAACAAAATGACGGGGAAAGATGTTGCGGGACGCACAAAATTGCC
>de3-416001-416101	CCCGATAAATCCGGCACCAAACAACCAGGTCAACGCCAGCCAGGAGATAACCTGGCTT**TTGTTGTTTTTGTACAT**GGCG ATAGCCGCCATGCCGTAGGTGA
>de3-485501-485601	ATCTCATTGATTCGTAACCTGCAATTTATGCCGAAGGTTTTATTGCTGGATGAAATAACCAGTGCGCTGG**ATGAAAGTA ATAAACAT**AACGTCAATGAGAT
>de3-894501-894601	AATTCGGCGACGTTTATATTCACGGAGGCTAACTTCTTGTTTACTAGTAAGCTGTTGAAAGAA**ATGGTAATTTACGATA A**TGTTTTTTACCAGAATTCAGG
>de3-1474901-1475001	AAAGCAAAAGACAAAGAGGGATGCAGACATTGTCTTG**CTGAAATTATGCAACAA**AATATTGCTATTTTATACCAGCAAT ACAATCGTTAAAGTAAATGCCC
>de3-1882101-1882301	GGTGCACTGTACTGCTTTTACGAGCTTGCGAAAACTGTAAACGCTTATCCACCCGTGCGATTACGGGAAAAGCGCGCAA AGTGCGGCAAAACTGATAAAAAAATCATCGTTTTCCATTCGGTAAAAATCTGACACTGATCATGTT**ATGAAAAAAAATA ACAA**CTTTTTTATCTGCTTTTGTCATTAACACGGCACACAGGC
>de3-1944351-1944451	CGATTACCGTGTCGTATTTAAAACCGGGGACTATTTCAAGAAACAAAATC**TTGAAAGTTTCTTCCCT**GAGATCCCCGTT GAATTTCATATTAATAAAGTGA
>de3-2591701-2591801	GACGTAACCACAACGCCACAACAGCTAACACAGCTCCTAACG**CGAAAGGAATACGCCAT**CCCCACTCTCTGAGTGCAGC GTCTTCCATGGTGTGTTGTAAA
>de3-2781001-2781151	TTATCGGAATGCGTGTTCTGGTGAACTTTTGGCTTACGGTTGTGATGTTGTGTTGTTGTGTTTGCAATTGGTCTGCGAT TCAGACCA**TGGTAGCAAAGCTACCT**TTTTTCACTTCCTGTACATTTACCCTGTCTGTCCATAGTGATTAATG
>de3-3114051-3114201	CCGATTCACCAGGAATAATGAATGGAACTTTTGACCCAATTGCTGCAAGCCCTGTGGGCGCAGGATTTTGAAACCCTGG CCAATCCATCGATGATTGGCATGTTGTATTTTGTCTTGT**TTGTAATTTTGTTCCTT**GAAAACGGCTTGCTTC
>de3-3289401-3289501	TACCGCGGCTGCTGGCACGGAGTTAGCCGGTGCTTCTTCTGCGGGTAACGTCAATGAGCAAAGG**TATTAACTTTACTCC CT**TCCTCCCCGCTGAAAGTACT
>de3-3290251-3290401	CAAGTATTTTTTTGATCTTTTTTTCTGT**TTGTTGTTTTTTCACCC**TTTTTGCTGCAATCGCACACAAAACGGTGCTTTT TTGCATACTAAAAGACTTGCACAAGGCCAATAATGCCCCCAAAGTCATTAGTAAATCATTTATTGCTGAGGT
>de3-3311201-3311301	TCGTCAATGTCAGCGCCATCGTTGTGTTCAGCGTTAGCAATGGCAGATTCCAGAACTTTCTTGACCAGTACAGCCGCTT TC**TTGTTGGTGTAGGTCAA**AAT
>de3-3593101-3593201	ATAAGTACAATTGCGCAACAAAAGTAAGATCTCGGTCATAAATCAAGAAATAAACCAAAAATCGTAATCGAAAGATAAA AAT**CTGTAATTGTTTTCCCC**TG
>de3-3635101-3635201	CACAAAGATTCAACAAACCATCAAAACAAAAATGTGACACTACTCACATTTAAATGCCATTTTTAGCGAAAATCGCCGC C**TTGTTGCTTTTTTACAC**AAGC
>de3-3654351-3654501	GCATCGGCTTGCCGCGAGCGGTCGTATATTTTGATCTTCAGAGGCTATTTTATCGATTCAG**CTGTAGTAAAATTACGC**A AATTTTGACTCTTGAGTATGAGGTTGTCGCAATGTCGGTTTCTAAAAAACCTATGGTACTGGTGATTCTGGA
>de3-3671701-3671851	CTAAACGCTGGATATCGGCTTCAGGGATATCATCAATAAAAACGTGGAAAACAAAACTCACG**TCGTTGTTATGTAACAA** AACTGACGTGATTGAAACACCACAACCAAAAAGAAAGTTTTTATCAATGCCATAAGCAACATGGAATGATGA
>de3-3818501-3818651	CTGAATAAAGCGTGTTGCTTACGCCAGTCGCTTTCAAGCTCTTCCAG**TTGTTGTTTTATTTCAC**CTGGCAGACCTTCAG CCGATGAACCGGGAAGTGCCAGCGTACTGCCTTGTAAACTTACATCGCGTACAACAAGATGTTGGGCGCTAT
>de3-3956501-3956651	ACATGCCTTTTTTCGTAAGTAAGCAACATAAGCTGTCACGTTTTGTGATGGCTATTAGAAATTCCTATGCAACAA**CTGA AAAAAAATTACAA**AAAGTGCTTTCTGAACTGAACAAAAAAGAGTAAAGTTAGTCGCGTAGGGTACAGAGGTA
>de3-3980201-3980351	GACATGCTTTGCTTGTCTGTTTTTGATCGTAT**TTGTAATTTATCGTCAA**AAAATTGACAGCCGTCACTTTTTAAACAAT TGGTGAAATTAAAATGAACGCATCCCAAAATGTTTAAGGAATGACCATGATTCGTGTTGCTTGTGTAGGTAT
>de3-4316551-4316651	CCAGATTTGGATGGTTTTTTTTTGTCTATAGC**TGGTAAGATAATTACGT**ACTGCAAATATTCATTTTATTAATATTTAA ACTTAATCTATATGGAATAATA
>de3-4413751-4413901	GATCCTTAACTGATCGGCATTAATCTTGGTTCTGGTGT**TTGTAACAAACTATCAG**CTACAAAAATATGCTCAATTTGTG ACATCAGTAACAAAACGCGTTTTGTTATGTGGATTGCTTGTTTTTTGATCTGGTATAACAGGTATAAAGGTA
>de3-426601-426901	GAATGGCCACCTGTAAGCTCCAGATGACCATTTTTGTTATTCTCCACAACGAGTTAGTTCTTCTTTTCGGATCCGGCAC TTCTGGGGGGGAAATCCAGCGATGGCTGGATTATGTCGTCAATTAAAAATGCGGCGAGTAGATTAGCAAATATCCACGC TTTCGCGAGTTCAGGTTCCTTTGCACGCAAAGCATCCAGGTGCAGCAAACTTTTGAGCCGCTTAAAAGCCAGTTCAATT TGCCATCGCAGACGGTAACAATCAGCCACTTGCTCTGCTGAATATTCATCTTCCGGTAATGATG
>de3-1589301-1589401	AACCCAACAATCAGATCGCCCATGCTTACGTGCAAGCGCCATAAAAGTCATGCCGCCGACATAATAGTCCACCAATAAA TCATGTAAATCGCTGTTGTTTT
>de3-3823151-3823401	GCGGTATCTCCGCAGCCCTGAGAAGCTCATCAAGAGGATAAAACTGCCTTAGCTCGTTGAGTACTTTCACTTTTTCGTG GGATGAGCTAAGGCTTTCAGCTTTTTTAGATACATAAGCCGCGTTTCAAGAAATCGAACTTGCCTTTCAAGATCCTCAA TGCGTCGGTCTTTTGACAGCTCCAATGCTGATGCCGCTTTTTCTGGATCAACTGATATTGCAATGTTTCTTTTGGTGCC AATCTTGAGCGCGC

Appendix III

List of HexR-like proteins in different *Enterobacterial* spciecs. Summary of this data is described in Chapter 5.

Transcription Factor	Pathway	Co-factor	% identity with YebK (Protein)
AguR	Agmatine utilization		30%
AlsR	Allose utilization	D-allose	25%
GlvR Staphylococcaceae	Maltose utilization	Maltose-6-Phosphate	23%
GntR1 - Lactobacillaceae	Gluconate utilization	Gluconate	31%
HexR – Oceanospirillales/ Alteromonadales	Central Carbohydrate Metabolism	KDPG	53%
HexR - Burkholderia	Central Carbohydrate Metabolism	KDPG	40%
HexR - Comamonadaceae	Central Carbohydrate Metabolism	KDPG	38%
HexR - Ralstonia	Central Carbohydrate Metabolism	KDPG	37%
HexR – Various betaproteobacteria	Central Carbohydrate Metabolism	KDPG	43%
HexR - Alteromonadales	Central Carbohydrate Metabolism	KDPG	71%
HexR - Pseudomonadaceae	Central Carbohydrate Metabolism	KDPG	58%
HexR2 – Oceanospirillales/ Alteromonadales	Central Carbohydrate Metabolism	KDPG	57%
MurR - Bacillales	N-acetylmuramate utilization	N-acetylmuramate-6-phosphate	31%
NanR - Staphylococcaceae	Sialic acid utilization	N-acetylmannosamine-6-phosphate	23%
NanR - Streptococcaceae	Sialic acid utilization	N-acetylneuraminic acid	21%
RbtR - Lactobacillaceae	Ribitol utilization		25%
SiaR - Enterobacteriales	Sialic acid utilization	Glucosamine-6-phosphate	28%
TagR - Lactobacillaceae	Tagatose utilization	Tagatose-6-phosphate	24%
TagR - Streptococcaceae	Tagatose utilization	Tagatose-6-phosphate	26%
UctR - Thermotogales	Sugar utilization		31%

Acknowledgments

Mom, Dad, Teacher and the God: An old Indian saying goes. I would like to extend my profound thanks to my mom, dad, siblings and niblings for their esteemed support and encouragement throughout my PhD. Their love, blessings, care and prayers always remained as a best part of my life even when miles kept as away. Then I would like to thank Prof. Sung Kuk Lee for his support, encouragement and kindness he shared for the past 4 years of my stay with him. Though we shared a cup of hard times, a sign of kindness in his eyes remained as a source of encouragement to do good science and taught me that even failures have a sweet taste. I would also like to thank Prof. Robert James Mitchel and Prof. Cheol-Min Ghim for continuously fuelling my passion and interest in science. I would like to specially thank Prof. Cheol-Min Ghim for serving as a role model for a better professional career and refining me to orient towards a global culture. I would like to thank my PhD committee members, Prof. Dougu Nam and Prof. Gyoo Yeol Jung for their support in science. Last in turn I would like to thank the Almighty for always keeping me warm through all leaps and falls of my PhD and giving me a confidence for life whenever I fail.

I would also like to thank my best friend, Ms. Gomathi, for being with me in all times of my life. I would extend my grateful thanks to my senior Mr. Jae Myung Lee who always made me feel-at-home even on a foreign land and extend his caring hands. I would also like to thank my lab seniors, Ms. Jung Min Park, Ms. Jihye Park, Ms. Goo Hee Kim, Dr. Sang Kyu Jang, Ms. Min-Kyung Kim, Dr. Su-Min Kim, Ms. Azar, Dr. Rajesh, Mr. Kwangsu Shin and Ms. Euna Lee for their kindness and support. The smile on their face and greetings meant a lot to my lonely life. I would like to thank Mr. Baeyoung Choi for being a true friend of mine. I would like to thank the active and smart undergraduates of my lab, Mr. Baeyoung Choi, Mr. Yong Sun Park, Ms. Ee Seul Shin, Ms. Yulim Lee, Mr, Jaesung Yoo, Mr. Su Min and Ms. Jee Won Choi for being a friendly companion of mine without any age disparity.

I would also like to thank my off lab friends, Dr. Vijaya, Dr. Mohammed uncle, Ms. Ghana, Mr. Vijay, Dr. Joel, Dr. Palani and their family for making me feel the wrath of a family and shared the love with me.

I would like to thank the incubators and HPLC in my lab for being my companion even during lonely long holidays and talking to me in a sweet language. Here I learnt several new things for the first time including cooking food, climbing a ladder, mountain climbing, ski-camp, karaoke singing, non-vegetarian diet, etc. Each one was a new experience to keep in mind.

I would like to thank Ms. Young Shin Ryu, Dr. Sang Woo Kim, Mr. Jiwon Lim, and Dr. Sathesh for helping me to understand the politics in science and teaching me tough lessons of my life. Of last, I would like to thank all members of my lab for their support and encouragement throughout my PhD.

Finally I would like to dedicate my PhD work to my ***E. coli*** (my love) for fuelling the passion of science in me, staying as a best enjoyable companion of me and for educating me to be refined and I extend my happiness that my PhD defense is on the birthday of my strain, OSS. Thank you *E. coli* for making me what I am today.

References

1. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H., Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2006,** *2*.
2. Khan, M. A.; Isaacson, R. E., *In vivo* expression of the β-glucoside (*bgl*) operon of *Escherichia coli* occurs in mouse liver. *Journal of Bacteriology* **1998,** *180* (17), 4746-4749.
3. Madan, R.; Kolter, R.; Mahadevan, S., Mutations that activate the silent *bgl* operon of *Escherichia coli* confer a growth advantage in stationary phase. *Journal of Bacteriology* **2005,** *187* (23), 7912-7917.
4. An, C. L.; Lim, W. J.; Hong, S. Y.; Shin, E. C.; Kim, M. K.; Lee, J. R.; Park, S. R.; Woo, J. G.; Lim, Y. P.; Yun, H. D., Structural and biochemical analysis of the *asc* operon encoding 6-phospho-β-glucosidase in *Pectobacterium carotovorum* subsp. carotovorum LY34. *Research in Microbiology* **2005,** *156* (2), 145-153.
5. Chen, L.; Vitkup, D., Distribution of orphan metabolic activities. *Trends in Biotechnology* **2007,** *25* (8), 343-348.
6. Denger, K.; Weiss, M.; Felux, A.-K.; Schneider, A.; Mayer, C.; Spiteller, D.; Huhn, T.; Cook, A. M.; Schleheck, D., Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature* **2014,** *507* (7490), 114-117.
7. Oleksiuk, O.; Jakovljevic, V.; Vladimirov, N.; Carvalho, R.; Paster, E.; Ryu, William S.; Meir, Y.; Wingreen, Ned S.; Kollmann, M.; Sourjik, V., Thermal robustness of signaling in bacterial chemotaxis. *Cell* **2011,** *145* (2), 312-321.
8. Saito, N.; Robert, M.; Kochi, H.; Matsuo, G.; Kakazu, Y.; Soga, T.; Tomita, M., Metabolite profiling reveals YihU as a novel hydroxybutyrate dehydrogenase for alternative succinic semialdehyde metabolism in *Escherichia coli*. *Journal of Biological Chemistry* **2009,** *284* (24), 16442-16451.
9. Kachroo, A. H.; Kancherla, A. K.; Singh, N. S.; Varshney, U.; Mahadevan, S., Mutations that alter the regulation of the *chb* operon of *Escherichia coli* allow utilization of cellobiose. *Molecular Microbiology* **2007,** *66*, 1382-1395.
10. Kricker, M.; Hall, B. G., Directed evolution of cellobiose utilization in *Escherichia coli* K12. *Molecular Biology and Evolution* **1984,** *1* (2), 171-182.
11. Neelakanta, G.; Sankar, T. S.; Schnetz, K., Characterization of a β-glucoside operon (*bgc*) prevalent in septicemic and uropathogenic *Escherichia coli* strains. *Applied and Environmental Microbiology* **2009,** *75* (8), 2284-2293.
12. Keyhani, N. O.; Roseman, S., Wild-type *Escherichia coli* grows on the chitin disaccharide, N,N′-diacetylchitobiose, by expressing the cel operon. *Proceedings of the National Academy of Sciences* **1997,** *94* (26), 14367-14371.
13. Plumbridge, J.; Pellegrini, O., Expression of the chitobiose operon of *Escherichia coli* is regulated by three transcription factors: NagC, ChbR and CAP. *Molecular Microbiology* **2004,** *52* (2), 437-449.
14. Thompson, J.; Ruvinov, S. B.; Freedberg, D. I.; Hall, B. G., Cellobiose-6-phosphate hydrolase (CelF) of *Escherichia coli:* Characterization and assignment to the unusual family 4 of glycosylhydrolases. *Journal of Bacteriology* **1999,** *181* (23), 7339-7345.
15. Hall, B. G.; Xu, L., Nucleotide sequence, function, activation, and evolution of the cryptic *asc* operon of *Escherichia coli* K12. *Molecular Biology and Evolution* **1992,** *9* (4), 688-706.
16. Mukerji, M.; Mahadevan, S., Cryptic genes: evolutionary puzzles. *Journal of Genetics* **1997,** *76* (2), 147-159.
17. Ishida, Y.; Kori, A.; Ishihama, A., Participation of regulator AscG of the β-glucoside utilization operon in regulation of the propionate catabolism operon. *J. Bacteriol.* **2009,** *191* (19), 6136-6144.
18. Vinuselvi, P.; Lee, S. K., Engineering *Escherichia coli* for efficient cellobiose utilization. *Appl Microbiol Biotechnol* **2011,** *92* (1), 125-32.
19. Alper, H.; Stephanopoulos, G., Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nature reviews. Microbiology* **2009,** *7* (10), 715-723.
20. Alper, H.; Fischer, C.; Nevoigt, E.; Stephanopoulos, G., Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (36), 12678-12683.
21. Chu, S.; Majumdar, A., Opportunities and challenges for a sustainable energy future. *Nature* **2012,** *488* (7411), 294-303.
22. Lee, S. K.; Chou, H.; Ham, T. S.; Lee, T. S.; Keasling, J. D., Metabolic engineering of microorganisms

for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology* **2008,** *19* (6), 556-563.
23. Li, H.; Cann, A. F.; Liao, J. C., Biofuels: biomolecular engineering fundamentals and advances. *Annual Review of Chemical and Biomolecular Engineering* **2010,** *1* (1), 19-36.
24. Zaldivar, J.; Nielsen, J.; Olsson, L., Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Applied Microbiology and Biotechnology* **2001,** *56* (1-2), 17-34.
25. Zhang, Y.-H. P.; Lynd, L. R., Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (20), 7321-7325.
26. Vinuselvi, P.; Lee, S. K., Engineering *Escherichia coli* for efficient cellobiose utilization. *Applied Microbiology and Biotechnology* **2011,** *92* (1), 125-132.
27. Schellenberger, S.; Drake, H. L.; Kolb, S., Functionally redundant cellobiose-degrading soil bacteria respond differentially to oxygen. *Applied and Environmental Microbiology* **2011,** *77* (17), 6043-6048.
28. Teugjas, H.; Valjamae, P., Product inhibition of cellulases studied with 14C-labeled cellulose substrates. *Biotechnology for Biofuels* **2013,** *6* (1), 104.
29. Xiao, Z. Z.; Zhang, X.; Gregg, D. J.; Saddler, J. N., Effects of sugar inhibition on cellulases and beta-glucosidase during enzymatic hydrolysis of softwood substrates. *Applied Biochemistry and Biotechnology* **2004,** *113*, 1115-1126.
30. Ha, S.-J.; Galazka, J. M.; Rin Kim, S.; Choi, J.-H.; Yang, X.; Seo, J.-H.; Louise Glass, N.; Cate, J. H. D.; Jin, Y.-S., Engineered *Saccharomyces cerevisiae* capable of simultaneous cellobiose and xylose fermentation. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108* (2), 504-509.
31. (a) Xu, Q.; Singh, A.; Himmel, M. E., Perspectives and new directions for the production of bioethanol using consolidated bioprocessing of lignocellulose. *Curr Opin Biotechnol* **2009,** *20* (3), 364-371; (b) Panesar, P. S.; Marwaha, S. S.; Kennedy, J. F., *Zymomonas mobilis*: an alternative ethanol producer. *J Chem Technol Biotechnol* **2006,** *81* (4), 623-635; (c) Atsumi, S.; Cann, A. F.; Connor, M. R.; Shen, C. R.; Smith, K. M.; Brynildsen, M. P.; Chou, K. J. Y.; Hanai, T.; Liao, J. C., Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metabolic Engineering* **2008,** *10* (6), 305-311; (d) Peralta-Yahya, P. P.; Keasling, J. D., Advanced biofuel production in microbes. *Biotechnol J* **2010,** *5* (2), 147-162.
32. Clomburg, J. M.; Gonzalez, R., Biofuel production in *Escherichia coli*: the role of metabolic engineering and synthetic biology. *Appl. Microbiol. Biotechnol.* **2010,** *86* (2), 419-434.
33. Hall, B. G.; Betts, P. W., Cryptic genes for cellobiose utilization in natural isolates of *Escherichia coli*. *Genetics* **1987,** *115* (3), 431-439.
34. Hall, B. G.; Betts, P. W.; Kricker, M., Maintenance of the cellobiose utilization genes of *Escherichia coli* in a cryptic state. *Mol Biol Evol* **1986,** *3* (5), 389-402.
35. Edwards, M. C.; Henriksen, E. D.; Yomano, L. P.; Gardner, B. C.; Sharma, L. N.; Ingram, L. O.; Doran Peterson, J., Addition of genes for cellobiase and pectinolytic activity in *Escherichia coli* for fuel ethanol production from pectin-rich lignocellulosic biomass. *Appl Environ Microbiol* **2011,** *77* (15), 5184-5191.
36. Sekar, R.; Shin, H.-D.; Chen, R., Engineering *Escherichia coli* cells for cellobiose assimilation through a phosphorolytic mechanism. *Applied and Environmental Microbiology* **2012,** *78* (5), 1611-1614.
37. Blattner, F. R.; Plunkett, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y., The complete genome sequence of *Escherichia coli* K-12. *Science* **1997,** *277* (5331), 1453-1462.
38. Datsenko, K. A.; Wanner, B. L., One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of National Academy of Sciences of the United States of America* **2000,** *97* (12), 6640-6645.
39. Cherepanov, P. P.; Wackernagel, W., Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of FLP-catalyzed excision of the antibiotic-resistance determinant. *Gene* **1995,** *158* (1), 9-14.
40. Jensen, P. R.; Hammer, K., The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl. Environ. Microbiol.* **1998,** *64* (1), 82-87.
41. Adin, D. M.; Visick, K. L.; Stabb, E. V., Identification of a cellobiose utilization gene cluster with cryptic beta-galactosidase activity in *Vibrio fischeri*. *Appl. Environ. Microbiol.* **2008,** *74* (13), 4059-4069.
42. Hall, B. G.; Xu, L., Nucleotide- sequence, function, activation and evolution of the cryptic *asc* operon of *Escherichia coli* K- 12. *Mol. Biol. Evol.* **1992,** *9* (4), 688-706.
43. (a) La Grange, D.; den Haan, R.; van Zyl, W., Engineering cellulolytic ability into bioprocessing organisms. *Appl. Microbiol. Biotechnol.* **2010,** *87* (4), 1195-1208; (b) Moniruzzaman, M.; Lai, X.; York, S.; Ingram, L., Isolation and molecular characterization of high-performance cellobiose- fermenting spontaneous mutants of ethanologenic *Escherichia coli* KO11 containing the *Klebsiella oxytoca casAB* operon. *Applied and*

Environmental Microbiology **1997,** *63* (12), 4633-4637.
44. Negrete, A.; Ng, W.-I.; Shiloach, J., Glucose uptake regulation in *E. coli* by the small RNA SgrS: comparative analysis of *E. coli* K-12 (JM109 and MG1655) and *E. coli* B (BL21). *Microb Cell Fact* **2010,** *9* (1), 75.
45. (a) van Rooyen, R.; Hahn-Hägerdal, B.; La Grange, D. C.; van Zyl, W. H., Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. *J. Biotechnol.* **2005,** *120* (3), 284-295; (b) Lee, S. M.; Jin, L. H.; Kim, J. H.; Han, S. O.; Na, H. B.; Hyeon, T.; Koo, Y. M.; Kim, J.; Lee, J. H., Beta-Glucosidase coating on polymer nanofibers for improved cellulosic ethanol production. *Bioprocess and Biosystems Engineering.* **2010,** *33* (1), 141-147.
46. Park, J. H.; Lee, K. H.; Kim, T. Y.; Lee, S. Y., Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proceedings of the National Academy of Sciences* **2007,** *104* (19), 7797-7802.
47. Horinouchi, T.; Tamaoka, K.; Furusawa, C.; Ono, N.; Suzuki, S.; Hirasawa, T.; Yomo, T.; Shimizu, H., Transcriptome analysis of parallel-evolved *Escherichia coli* strains under ethanol stress. *BMC Genomics* **2010,** *11* (1), 579.
48. Orencio-Trejo, M.; Flores, N.; Escalante, A.; Hernandez-Chavez, G.; Bolivar, F.; Gosset, G.; Martinez, A., Metabolic regulation analysis of an ethanologenic *Escherichia coli* strain based on RT-PCR and enzymatic activities. *Biotechnology for Biofuels* **2008,** *1* (1), 8.
49. Gonzalez, R.; Tao, H.; Shanmugam, K. T.; York, S. W.; Ingram, L. O., Global gene expression differences associated with changes in glycolytic flux and growth rate in *Escherichia coli* during the fermentation of glucose and xylose. *Biotechnology Progress* **2002,** *18* (1), 6-20.
50. Troein, C.; Ahrén, D.; Krogh, M.; Peterson, C., Is transcriptional regulation of metabolic pathways an optimal strategy for fitness? *PLoS ONE* **2007,** *2* (9), e855.
51. Datta, S.; Costantino, N.; Court, D. L., A set of recombineering plasmids for gram-negative bacteria. *Gene* **2006,** *379*, 109-115.
52. Lee, S. K.; Newman, J. D.; Keasling, J. D., Catabolite repression of the propionate catabolic genes in *Escherichia coli* and *Salmonella enterica*: evidence for involvement of the cyclic AMP receptor protein. *Journal of Bacteriology* **2005,** *187* (8), 2793-2800.
53. (a) Oh, M.-K.; Rohlin, L.; Kao, K. C.; Liao, J. C., Global expression profiling of acetate-grown *Escherichia coli*. *Journal of Biological Chemistry* **2002,** *277* (15), 13175-13183; (b) Oh, M.-K.; Liao, J. C., Gene Expression Profiling by DNA Microarrays and Metabolic Fluxes in *Escherichia coli*. *Biotechnology Progress* **2000,** *16* (2), 278-286.
54. Chi, S.-M.; Kim, J.; Kim, S.-Y.; Nam, D., ADGO 2.0: interpreting microarray data and list of genes using composite annotations. *Nucleic Acids Research* **2011,** *39* (suppl 2), W302-W306.
55. Feist, A. M.; Henry, C. S.; Reed, J. L.; Krummenacker, M.; Joyce, A. R.; Karp, P. D.; Broadbelt, L. J.; Hatzimanikatis, V.; Palsson, B. O., A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology* **2007,** *3*.
56. Li, M. Z.; Elledge, S. J., Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* **2007,** *4* (3), 251-256.
57. Moen, B.; Janbu, A. O.; Langsrud, S.; Langsrud, Ø .; Hobman, J. L.; Constantinidou, C.; Kohler, A.; Rudi, K., Global responses of *Escherichia coli* to adverse conditions determined by microarrays and FT-IR spectroscopy. *Canadian Journal of Microbiology* **2009,** *55* (6), 714-728.
58. Zhao, K.; Liu, M.; Burgess, R. R., Adaptation in bacterial flagellar and motility systems: from regulon members to 'foraging'-like behavior in *E. coli*. *Nucleic Acids Research* **2007,** *35* (13), 4441-4452.
59. Jozefczuk, S.; Klie, S.; Catchpole, G.; Szymanski, J.; Cuadros-Inostroza, A.; Steinhauser, D.; Selbig, J.; Willmitzer, L., Metabolomic and transcriptomic stress response of *Escherichia coli*. *Molecular Systems Biology* **2010,** *6*.
60. Franchini, A. G.; Egli, T., Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. *Microbiology* **2006,** *152* (7), 2111-2127.
61. Kachroo, A. H.; Kancherla, A. K.; Singh, N. S.; Varshney, U.; Mahadevan, S., Mutations that alter the regulation of the *chb* operon of *Escherichia coli* allow utilization of cellobiose. *Molecular Microbiology* **2007,** *66* (6), 1382-1395.
62. Datta, S.; Costantino, N.; Court, D. L., A set of recombineering plasmids for gram-negative bacteria. *Gene* **2006,** *379* (0), 109-115.
63. Datsenko, K. A.; Wanner, B. L., One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences* **2000,** *97* (12), 6640-6645.

64. Alqasim, A.; Emes, R.; Clark, G.; Newcombe, J.; La Ragione, R.; McNally, A., Phenotypic microarrays suggest *Escherichia coli* ST131 is not a metabolically distinct lineage of extra-intestinal pathogenic *E. coli*. *PLoS ONE* **2014,** *9* (2), e88374.
65. Ryu, Y. S.; Biswas, R. K.; Shin, K.; Parisutham, V.; Kim, S. M.; Lee, S. K., A simple and effective method for construction of *Escherichia coli* strains proficient for genome engineering. *PLoS ONE* **2014,** *9* (4), e94266.
66. Na, D.; Lee, D., RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinformatics* **2010,** *26* (20), 2633-2634.
67. Scotto-Lavino, E.; Du, G.; Frohman, M. A., 5' end cDNA amplification using classic RACE. *Nat. Protocols* **2007,** *1* (6), 2555-2562.
68. Daddaoua, A.; Krell, T.; Ramos, J.-L., Regulation of glucose metabolism in *Pseudomonas*: the phosphorylative branch and Entner-Doudoroff enzymes are regulated by a repressor contatining a sugar isomerase domain. *Journal of Biological Chemistry* **2009,** *284* (32), 21360-21368.
69. Thompson, J.; Ruvinov, S. B.; Freedberg, D. I.; Hall, B. G., Cellobiose-6-phosphate hydrolase (CelF) of *Escherichia coli*: characterization and assignment to the unusual family 4 of glycosylhydrolases. *Journal of Bacteriology* **1999,** *181* (23), 7339-7345.
70. Parisutham, V.; Jung, S.-K.; Nam, D.; Lee, S. K., Transcriptome-driven synthetic re-modeling of *Escherichia coli* to enhance cellobiose utilization. *Chemical Engineering Science* **2013,** *103* (0), 50-57.
71. Raghavan, R.; Sage, A.; Ochman, H., Genome-wide identification of Transcription Start Sites yields a novel thermosensing RNA and new cyclic AMP receptor protein-regulated genes in *Escherichia coli*. *Journal of Bacteriology* **2011,** *193* (11), 2871-2874.
72. Rolfe, M. D.; Rice, C. J.; Lucchini, S.; Pin, C.; Thompson, A.; Cameron, A. D. S.; Alston, M.; Stringer, M. F.; Betts, R. P.; Baranyi, J.; Peck, M. W.; Hinton, J. C. D., Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *Journal of Bacteriology* **2012,** *194* (3), 686-701.
73. Ito, M.; Baba, T.; Mori, H.; Mori, H., Functional analysis of 1440 *Escherichia coli* genes using the combination of knock-out library and phenotype microarrays. *Metabolic Engineering* **2005,** *7* (4), 318-327.
74. Kochanowski, K.; Sauer, U.; Chubukov, V., Somewhat in control: the role of transcription in regulating microbial metabolic fluxes. *Current Opinion in Biotechnology* **2013,** *24* (6), 987-993.
75. (a) Cheng, K.-K.; Lee, B.-S.; Masuda, T.; Ito, T.; Ikeda, K.; Hirayama, A.; Deng, L.; Dong, J.; Shimizu, K.; Soga, T.; Tomita, M.; Palsson, B. O.; Robert, M., Global metabolic network reorganization by adaptive mutations allows fast growth of *Escherichia coli* on glycerol. *Nat Commun* **2014,** *5*; (b) Hua, Q.; Joyce, A. R.; Palsson, B. O.; Fong, S. S., Metabolic characterization of *Escherichia coli* strains adapted to growth on lactate. *Applied and Environmental Microbiology* **2007,** *73* (14), 4639-4647.
76. Conrad, T.; Joyce, A.; Applebee, M. K.; Barrett, C.; Xie, B.; Gao, Y.; Palsson, B., Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome biology* **2009,** *10* (10), R118.
77. (a) Jarboe, L. R.; Zhang, X. L.; Wang, X.; Moore, J. C.; Shanmugam, K. T.; Ingram, L. O., Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. *J. Biomed. Biotechnol.* **2010**; (b) Parisutham, V.; Jung, S. K.; Nam, D.; Lee, S. K., Transcriptome-driven synthetic re-modeling of *Escherichia coli* to enhance cellobiose utilization. *Chemical Engineering Science* **2013,** *103*, 50-57; (c) Eriksen, D.; Hsieh, P. C.; Lynn, P.; Zhao, H., Directed evolution of a cellobiose utilization pathway in *Saccharomyces cerevisiae* by simultaneously engineering multiple proteins. *Microbial Cell Factories* **2013,** *12* (1), 61.
78. Kotte, O.; Zaugg, J. B.; Heinemann, M., Bacterial adaptation through distributed sensing of metabolic fluxes. 2010; Vol. 6.
79. (a) Klumpp, S.; Hwa, T., Bacterial growth: global effects on gene expression, growth feedback and proteome partition. *Curr Opin Biotechnol* **2014,** *28C*, 96-102; (b) Xu, Y.-F.; Amador-Noguez, D.; Reaves, M. L.; Feng, X.-J.; Rabinowitz, J. D., Ultrasensitive regulation of anapleurosis via allosteric activation of PEP carboxylase. *Nat Chem Biol* **2012,** *8* (6), 562-568.
80. Dragosits, M.; Mattanovich, D., Adaptive laboratory evolution - principles and applications for biotechnology. *Microbial Cell Factories* **2013,** *12* (1), 64.
81. Sandberg, T. E.; Pedersen, M.; LaCroix, R. A.; Ebrahim, A.; Bonde, M.; Herrgard, M. J.; Palsson, B. O.; Sommer, M.; Feist, A. M., Evolution of *Escherichia coli* to 42°C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. *Molecular Biology and Evolution* **2014**.
82. (a) Hua, Q.; Joyce, A. R.; Palsson, B. Ø .; Fong, S. S., Metabolic characterization of *Escherichia coli* strains adapted to growth on lactate. *Applied and Environmental Microbiology* **2007,** *73* (14), 4639-4647; (b)

Johnson, M. D.; Bell, J.; Clarke, K.; Chandler, R.; Pathak, P.; Xia, Y.; Marshall, R. L.; Weinstock, G. M.; Loman, N. J.; Winn, P. J.; Lund, P. A., Characterization of mutations in the PAS domain of the EvgS sensor kinase selected by laboratory evolution for acid resistance in *Escherichia coli*. *Molecular Microbiology* **2014,** *93* (5), 911-927.
83. Leyn, S. A.; Li, X.; Zheng, Q.; Novichkov, P. S.; Reed, S.; Romine, M. F.; Fredrickson, J. K.; Yang, C.; Osterman, A. L.; Rodionov, D. A., Control of Proteobacterial central carbon metabolism by the HexR transcriptional regulator: a case study in *Shewanella oneidensis*. *Journal of Biological Chemistry* **2011,** *286* (41), 35782-35794.
84. Shimada, T.; Ishihama, A.; Busby, S. J. W.; Grainger, D. C., The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Research* **2008,** *36* (12), 3950-3955.
85. Wisselink, H. W.; Toirkens, M. J.; Wu, Q.; Pronk, J. T.; van Maris, A. J. A., Novel evolutionary engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by engineered *Saccharomyces cerevisiae* strains. *Applied and Environmental Microbiology* **2009,** *75* (4), 907-914.
86. (a) Yomano, L.; York, S.; Shanmugam, K.; Ingram, L., Deletion of methylglyoxal synthase gene (*mgsA*) increased sugar co-metabolism in ethanol-producing *Escherichia coli*. *Biotechnology Letters* **2009,** *31* (9), 1389-1398; (b) Vinuselvi, P.; Park, J. M.; Lee, J. M.; Oh, K.; Ghim, C.-M.; Lee, S. K., Engineering microorganisms for biofuel production. *Biofuels* **2011,** *2* (2), 153-166.
87. Lawford, H.; Rousseau, J., Relative rates of sugar utilization by an ethanologenic recombinant *Escherichia coli* using mixtures of glucose, mannose, and xylose. *Applied Biochemistry and Biotechnology* **1994,** *45-46* (1), 367-381.
88. Karimova, G.; Ladant, D.; Ullmann, A., Relief of catabolite repression in a cAMP-independent catabolite gene activator mutant of *Escherichia coli*. *Res. Microbiol.* **2004,** *155* (2), 76-79.
89. (a) Nair, N. U.; Zhao, H., Selective reduction of xylose to xylitol from a mixture of hemicellulosic sugars. *Metabolic Engineering* **2010,** *12* (5), 462-468; (b) Kimata, K.; Takahashi, H.; Inada, T.; Postma, P.; Aiba, H., cAMP receptor protein–cAMP plays a crucial role in glucose–lactose diauxie by activating the major glucose transporter gene in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **1997,** *94* (24), 12914-12919.
90. Inada, T.; Kimata, K.; Aiba, H. J., Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: challenge to the cAMP model. *Genes Cells* **1996,** *1* (3), 293-301.
91. Vinuselvi, P.; Lee, S., Engineering *Escherichia coli* for efficient cellobiose utilization. *Applied Microbiology and Biotechnology* **2011**, 1-8.
92. Nichols, N.; Dien, B.; Bothast, R., Use of catabolite repression mutants for fermentation of sugar mixtures to ethanol. *Applied Microbiology and Biotechnology* **2001,** *56*, 120 - 125.
93. Li, R.; Chen, Q.; Wang, P.; Qi, Q., A novel-designed *Escherichia coli* for the production of various polyhydroxyalkanoates from inexpensive substrate mixture. *Applied Microbiology and Biotechnology* **2007,** *75* (5), 1103-1109.
94. Yano, S.; Murakami, K.; Sawayama, S.; Imou, K.; Yokoyama, S., Ethanol production potential from oil palm empty fruit bunches in Southeast Asian countries considering xylose utilization. *Journal of the Japan Institute of Energy* **2009,** *88* (10), 923-926.
95. Sommer, M. O. A.; Church, G. M.; Dantas, G., A functional metagenomic approach for expanding the synthetic biology toolbox for biomass conversion. *Mol Syst Biol* **2010,** *6*.
96. (a) Graham, J. E.; Clark, M. E.; Nadler, D. C.; Huffer, S.; Chokhawala, H. A.; Rowland, S. E.; Blanch, H. W.; Clark, D. S.; Robb, F. T., Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat Commun* **2011,** *2*, 375; (b) Pottkamper, J.; Barthen, P.; Ilmberger, N.; Schwaneberg, U.; Schenk, A.; Schulte, M.; Ignatiev, N.; Streit, W. R., Applying metagenomics for the identification of bacterial cellulases that are stable in ionic liquids. *Green Chem* **2009,** *11* (7), 957-965.
97. Bokinsky, G.; Peralta-Yahya, P. P.; George, A.; Holmes, B. M.; Steen, E. J.; Dietrich, J.; Soon Lee, T.; Tullman-Ercek, D.; Voigt, C. A.; Simmons, B. A.; Keasling, J. D., Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proceedings of the National Academy of Sciences* **2011,** *108* (50), 19949-19954.
98. Qian, Z.-G.; Xia, X.-X.; Choi, J. H.; Lee, S. Y., Proteome-based identification of fusion partner for high-level extracellular production of recombinant proteins in *Escherichia coli*. *Biotechnol Bioeng* **2008,** *101* (3), 587-601.
99. Palmer, T.; Berks, B. C., The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Micro* **2012,** *10* (7), 483-496.
100. Ueda, M.; Tanaka, A., Cell surface engineering of yeast: construction of arming yeast with biocatalyst. *J Biosci Bioeng* **2000,** *90* (2), 125-136.